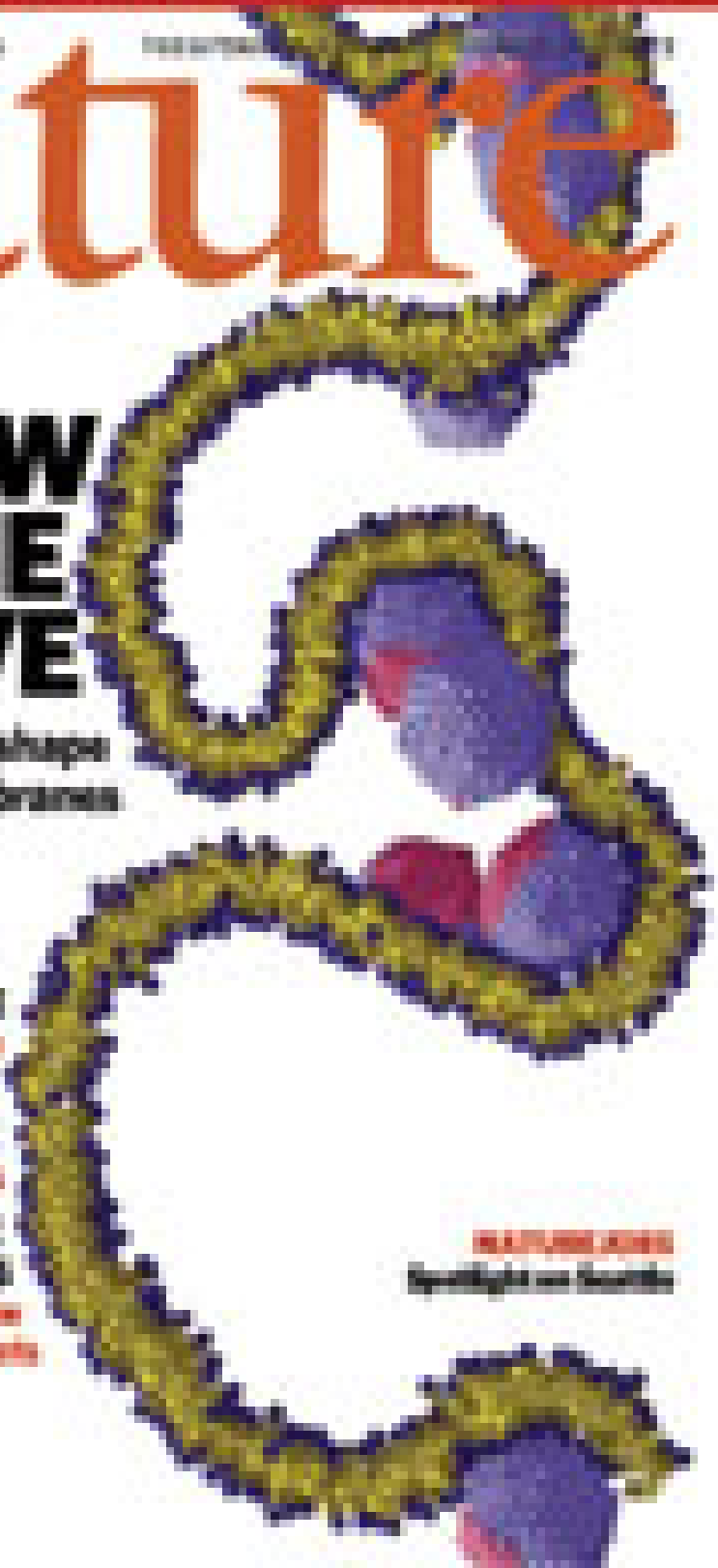
30 May [illegible] | www.nature.com/nature | £8

nature



FOLLOW THE CURVE

The forces that shape cell membranes

DOPAMINE JUBILEE
A molecule with a history

GLOBAL CLIMATE
The pull of the ocean

NEURODEGENERATIVE DISEASES
A clutch of new drug targets

NATURE JOBS
Spotlight on Seattle

nature

www.nature.com/nature

Vol 447 | Issue no. 7143 | 24 May 2007

Unwise branding

Equating animal-rights activism with terrorism increases the penalties for offenders and will please many of their victims. But it is not in the interests of science.

Terrorist is not a word you throw around lightly. And it is certainly not a word you apply to anyone with whom you would like to have a civil conversation. A US tendency to apply the label to militant activists who are against animal research or genetic engineering slams shut a door that might be difficult to reopen — to researchers' cost.

In a courtroom in Eugene, Oregon, last week, federal prosecutors asked for a 'terrorism enhancement' on the sentencing of ten environmental activists. The activists have admitted to a string of arson attacks in the western United States in the late 1990s and the start of this decade. They torched places where things were done of which they disapproved, including a lab that they believed was growing genetically engineered poplar trees. If the judge applies the requested enhancement, their sentences could be longer and the conditions of their imprisonment more severe.

They are criminals, to be sure. Their arson cost millions of dollars and destroyed scientific work in progress. But although some of their more knuckleheaded actions could easily have accidentally hurt someone, their ethos was to damage property, never to hurt or kill.

Other extreme activists are also being labelled terrorists. Last November, the Animal Enterprise Terrorism Act was signed into law in the United States. It creates tough penalties for damaging property, making threats and conspiring against zoos, animal labs and the like. Leaving aside the merits of this act, its very name enshrines into law the idea that destructive activists are terrorists.

As one of the communities targeted by these activists, scientists may be tempted to embrace this rhetoric. Indeed, many people have personally felt terrified by the actions of the most extreme. But 'terrorist' is a word so debased and loaded by political use that, if it has any meaning at all, it is counterproductive. There is no such objective thing as a terrorist. A criminal is a person who has been convicted of a crime. We can examine a person's records and make an unemotional determination of whether or not they are a criminal. But a terrorist is, in practice, a person who fights for a cause we do not believe in using methods that we do not approve of. Calling someone a terrorist is a value judgement.

It is a value judgement that seems to be increasingly used in the United States since the attacks of 11 September 2001. Indeed, the nation is waging, in official parlance, a "global war on terror". The term is useful politically exactly because it expresses an absolute rejection of a person and their aims. The terrorist label definitively ends any possibility of dialogue. But if there is any hope of bringing closer together those at the extremes of scientific controversies such as animal research and genetic engineering, the various parties must be able to speak to one another.

"We should avoid building an unbreachable wall between criminal activists and their victims."

Although most activists feel that the actions of the criminal few are unproductive and embarrassing, for every activist saboteur with a lighted match there are hundreds of people who are sympathetic to his or her cause. Label that saboteur a terrorist, and you risk alienating all of them. Efforts to bring together defenders and attackers of animal research, such as those by the UK-based Boyd Group, often do not admit those who espouse criminal acts, and that is appropriate. And it leaves open the possibility that an activist who has renounced criminal actions can come to the table. But who will be willing to publicly break bread with a terrorist, reformed or otherwise?

We should avoid building an unbreachable wall between criminal activists and their victims. The judge in this case should reject the call for 'terrorism enhancement'. We must all speak more objectively and calmly. ■

An unwieldy hybrid

A draft law will unnecessarily hinder embryo research.

The past few months have seen Britain's politicians tying themselves in knots over the question of whether to allow the creation of 'hybrid' embryos, those made from both human and animal material, for research purposes. Meanwhile, the embryologists who have applied for permission to carry out this research have waited patiently (or perhaps not so patiently) for the verdict.

The outlook initially looked bleak. In December, the government published a policy outline proposing a ban on virtually all forms of hybrid embryos. Medical research organizations reacted angrily, and the House of Commons Select Committee on Science and Technology began working on a report, unveiled in April, criticizing the proposed ban as unnecessary and unfairly restrictive.

Last week, the verdict arrived in the form of the government's draft bill, which seems to be a turnaround on the issue. If it becomes law, the new legislation looks set to allow the creation of 'cybrid embryos' — a particular form of hybrid in which human DNA is placed in an empty animal egg — by the two British research groups that have applied to do it. Assuming that the groups ultimately receive licences to create these embryos, they should give rise to valuable stem cells that could be used to study conditions such as Parkinson's disease.

A range of other techniques also look set to be approved, including the creation of human embryos with animal genes inserted in their DNA, human embryos containing animal cells, and genetically engineered animals with human genes (although the latter will now

fall under the purview of the government's animal-research guidelines). Researchers applying to create these specific entities look set to have their requests granted, within the existing rules that no *in vitro* human embryo should be allowed to develop beyond 14 days, and no embryo derived from animal material should be implanted in a human uterus.

In the main, research advocates are satisfied with the proposal, and the government is to be applauded for not persisting with its plan for an outright ban. But a closer reading of the draft bill reveals that the proposed legislation is prescriptive, in mind-boggling detail, rather than truly permissive — and this is an approach that looks set to harm the field of embryology in the longer term.

The bill states that the creation of hybrid embryos "should not be permitted but that there should be a regulation-making power allowing exceptions to the prohibition". That power will be the Human Fertilisation and Embryology Authority (HFEA, which is to be rebranded as the Regulatory Authority for Tissue and Embryos), and those exceptions will be the various strictly defined techniques prescribed in the draft act.

But politicians are not embryology experts, and in attempting to compose a definitive list of acceptable techniques, they risk saddling researchers with a piece of legislation that does not allow the freedom to pursue new and promising possibilities not covered by the draft bill. And no matter how thoroughly you consult stakeholders now, someone will come along in five or ten years' time and ask for permission to do something you hadn't thought of.

It was just such a situation that caused the recent ill-feeling over the government's handling of hybrid-embryo proposals. When separate research groups at Newcastle University and King's College London asked the HFEA last year for permission to create cybrid embryos, the agency panicked and referred the issue to the government, which proposed its infamous ban before later admitting that such research is necessary and useful. The proposed legislation will prompt a repeat performance every time researchers propose something that the regulatory body does not feel comfortable dealing with — and by inviting politicians into the fray with greater regularity, it encourages repeated attacks by those who want to see all work on human embryos outlawed.

Much better would be to preserve the spirit of Britain's 1990 Human Fertilisation and Embryology Act, which the draft bill is intended to replace. That act contained several general rules of thumb that have provided a useful ethical framework while not stifling research. Even 17 years down the line, those rules still stand up to scrutiny. Embryologists could look forward to a more fruitful future if they were given a regulatory body with the ethical muscle to approve novel techniques while adhering to tried and trusted principles. ■

Nobels in dubious causes

Top scientists should campaign only where they can truly make a difference.

With great power comes great responsibility, said the wise uncle of Peter Parker, a.k.a. Spiderman. The same might be true of Nobel laureates.

Every October, a new class of formerly obscure scientists is hurled into the limelight, their lives changing literally overnight with that phone call from Stockholm. Their daily routine changes from one of quiet hours in the lab to one encompassing many new demands on their time, from speaking engagements to invitations to sign the latest petition for peace and justice on the planet (see page 374).

In theory, this is a good thing. Most Nobel prizewinners are thoughtful people with insightful things to say about the world. And there is a rich history of prominent scientists playing crucial roles in major world decisions — Albert Einstein warning US President Franklin Roosevelt that the Germans might be thinking of building an atomic bomb, or the Federation of American Scientists drawing attention to the dangers of nuclear proliferation early in the atomic age.

But scientists need to take care not to overstep their expertise. It is reasonable to expect a Manhattan Project physicist to weigh in on the dangers of nuclear weapons, with which he or she is entirely familiar. It is less clear-cut to, say, support the candidacy of a politician.

In the United States, a group called Scientists and Engineers for America formed last year with the benevolent-sounding goals of good government, open debate, competent leadership and political participation. It sprang mainly, however, from years of frustration with the administration of President George W. Bush and its many instances of reportedly twisting science to its own ends. There is little doubt that US federal science has suffered under Bush, but it is unclear how this group will accomplish concrete goals to counter this.

Political advocacy can, in fact, be the trickiest road for a scientist–activist to navigate. Nobel-prizewinning economists, for instance, are routinely recruited to either side of US presidential campaigns, with their names trotted out like endorsements. In Scotland earlier this month, a group of 62 scientists (including Ian Wilmut, creator of Dolly the cloned sheep) wrote to *The Herald* newspaper, days before the country's elections, claiming that funding for science in Scotland would suffer in the event of "separation" from the United Kingdom. But the election wasn't about separation, it was about who was best equipped to run the Scottish parliament. The Scottish National Party won the election. In aligning themselves so clearly with the Labour Party's cack-handed attempts to scare its own former supporters back into the fold, the signatories at least ran the danger of seeming to be self-interested, grant-obsessed, and out of touch with people's desire for change.

Scientists who want to promote change in the world would be better off selecting their areas of activism carefully. Nobel laureates have a special responsibility, as they are regarded by the public with a level of awe. Many of them do use their names wisely to advance education or underappreciated areas of science. Last week, for instance, 40 of them helped launch a US$10-million fund to support scientific research in the Middle East. Such efforts are targeted, specific and worthy of the Nobel name. ■

RESEARCH HIGHLIGHTS

Cosmic rays peek inside

Nucl. Instrum. Methods Phys. Res. A **575,** 489–497 (2007)

Researchers in Japan have taken advantage of cosmic rays to image the inside of an active volcano. This approach has previously been used to search for chambers inside pyramids.

Hiroyuki Tanaka of the University of Tokyo and his colleagues placed an instrument that detects particles known as muons on the side of Mount Asama (pictured). Muons are sent off in all directions when cosmic rays hit Earth's atmosphere.

Some muons reach the detector having passed through the rocks of the volcano. By calculating the number of muons absorbed en route, the researchers determined the density of the volcano's innards. With more devices and real-time readings, the method may help in predicting eruptions.

KYODO NEWS/AP

METABOLISM

After-dinner protein

Cell **129,** 537–548 (2007)

Eat too much and one is at risk of obesity and its attendant chronic illnesses. Wrapped up in many of these illnesses is organ and tissue inflammation, which can be triggered by excess nutrients.

But surpluses are to be expected after meals, and so Gökhan Hotamisligil of the Harvard School of Public Health in Boston, Massachusetts, and his colleagues went looking for a mechanism that limits damage and inflammation at these times. A membrane-spanning protein known as STAMP2 fits the bill. It shows up in belly fat when mice eat. And mice with a STAMP2 deficiency had many of the symptoms of obesity-associated disease. The protein's precise function, though, remains a mystery.

ASTRONOMY

Old man of the heavens

Astrophys. J. **660,** L117–L120 (2007)

A mixture of radioactive isotopes has revealed a star in our Galaxy known as HE 1523-0901 to be almost as old as the Universe.

Anna Frebel of the University of Texas in Austin and her colleagues gathered data on the star's composition with the European Southern Observatory's Very Large Telescope in Chile. They calculated the star's age to be 13.2 billion years, give or take 2 billion years, by comparing the abundances of radioactive isotopes of uranium and thorium to each other and to other 'r-process' elements (the r-process happens in supernovae, forming certain heavy elements). The Universe is around 13.7 billion years old.

Studies of such ancient stars can provide hints of what happened to the first generation of short-lived stars, and what the Universe looked like chemically when it formed.

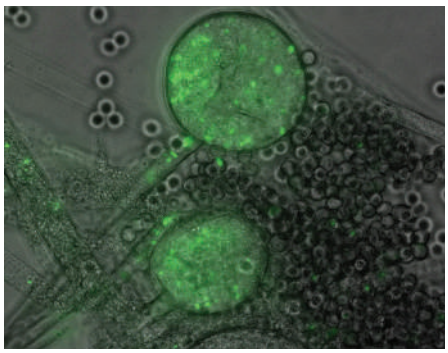
MICROBIOLOGY

Dependency in deadly duo

Curr. Biol. **17,** 773–777 (2007)

They are partners in crime. The fungus (*Rhizopus microsporus*) provides shelter to the bacterium (*Burkholderia*, shown green in image below) living in its cells. In turn, the bacterium produces a toxin that kills plants — supplying nutrients for the fungus. Together, they can wreak havoc on rice crops.

Now, Christian Hertweck of the Hans Knöll Institute in Jena, Germany, and his colleagues have found a new facet of the deadly duo's unusual bond — the fungus cannot reproduce asexually without the bacterium. The team observed that 'curing' *R. microsporus* of its bacterial partner by treating the fungus with antibiotics abolished spore formation, and that injecting labelled *Burkholderia* into cured *R. microsporus* cultures restored sporulation.



IMMUNOLOGY

Flies link sleep and sickness

Curr. Biol. **17,** R353–R355 (2007)

Sick flies don't sleep well, and flies that don't sleep well are susceptible to falling sick. This finding, from David Schneider of Stanford University in California and his colleagues, highlights a connection between circadian rhythm — an organism's daily clock — and immunity.

The researchers observed that *Drosophila* with bacterial infections lacked the daily pattern of movement seen in healthy flies. Also, flies with a mutation in either of two circadian genes died more quickly than controls when infected with bacteria. Infection may trigger a "spiral of death", says Schneider, in which worsening sleep worsens the illness. *Drosophila* could serve as a model organism for exploring the implications of the linkage for vertebrates, which share many circadian and immune signalling pathways with flies.

ELSEVIER

POLYMER PHYSICS

Forbidden symmetries

Phys. Rev. Lett. **98,** 195502 (2007)

Quasicrystals, usually associated with hard, inorganic materials such as metal alloys, have a softer side too. These materials have crystal-like structures with 'forbidden' symmetries that can't be produced by the regular packing of objects.

Kenichi Hayashida of Nagoya University in Japan and his co-workers report that a 'star' block copolymer, in which three different polymer chains are linked at a central focus, will form a type of quasicrystal. Like chains prefer to cluster together, producing segregation of chain types at the nanoscale. In a mixture with polystyrene, the segregation happens in a pattern with forbidden 12-fold symmetry.

The quasiperiodic spacing of 50 nanometres is five times greater than that seen for the only previous soft-matter quasicrystal — assembled from branched molecules known as dendrons — and is visible beneath an electron microscope.

METHODS

Tweezer pleaser

Nature Phys. doi:10.1038/nphys624 (2007)

Optical tweezers, which use light to grab and hold small particles or cells, have a new relation. Designed by Romain Quidant of the Institute of Photonic Sciences in Barcelona, Spain, the 'plasmonic' tweezer uses a single beam of light to arrange many small particles on a surface. Conventional optical tweezers would require a focused light beam to be deployed for each particle.

It works for particles placed on a glass slide decorated with dots of gold. When a beam of light shines on the slide, it excites in each dot a type of electromagnetic wave known as a surface plasmon. The plasmons concentrate light, trapping a single particle above each dot. Applied to a mixture of different-sized particles, the technique selectively traps those of a certain size. The hope is to integrate such systems into lab-on-a-chip devices.

CELL BIOLOGY

Torn to pieces

Dev. Cell **12,** 807–816 (2007)

The break-up of mitochondria, the powerhouses of a cell, seems to promote cell suicide in flies, report V. Sriram of the Tata Institute of Fundamental Research in Bangalore, India, and his colleagues.

Mitochondrial remodelling is important in guiding cells down the programmed cell-death pathway in worms and mammals, but it wasn't known whether the same was true for flies. Sriram's team observed in the fruitfly *Drosophila melanogaster* that mitochondria fragmentation precedes the initation of early steps in apoptosis. They also found evidence that Drp-1, a protein that is known to regulate mitochondrial fission in healthy cells, also regulates the break-up of mitochondria during cell death.

CANCER BIOLOGY

Destructive power

Science **316,** 1043–1046 (2007)

Researchers have discovered how mutations in the gene *WTX* drive cancerous growth in Wilms' tumours, a type of kidney cancer that mostly affects children.

The gene was first linked to the cancer in January. Using proteomic techniques, Randall Moon from the University of Washington in Seattle and his colleagues have now shown that the WTX protein is part of a complex that normally promotes destruction of β-catenin, a key signalling protein. β-catenin is involved in the WNT signal transduction pathway, which is important during embryo development and tissue repair.

Mutations in *WTX* lead to increased levels of β-catenin. This, in turn, promotes cell proliferation.

A. HENDRY

EVOLUTIONARY BIOLOGY

Size matters

Proc. R. Soc. Lond. B doi:10.1098/rspb.2007.0224 (2007)

The varied beaks of Darwin's finches are a famous example of speciation happening in geographically isolated populations. Now a team has found evidence, in the medium ground finch *Geospiza fortis* (pictured above), for what could be the early stages of sympatric speciation — in which populations evolve into different species without being physically separated.

Sarah Huber from the University of Massachusetts in Amherst and her colleagues have shown that two 'morphs' of the finch — one with a bigger beak than the other — prefer to mate with their own kind despite living side by side. Because beak size is related to song, females may select 'like males' on the basis of their song.

Furthermore, genetic analyses indicate reduced gene flow between the two populations.

JOURNAL CLUB

Peter S. Liss
University of East Anglia, Norwich, UK

A biogeochemist is keen to find out whether oceanic plankton can help to keep our planet cool.

For years I have been fascinated by the idea that oceanic plankton can play a significant part in controlling climate. This concept is, of course, at the heart of gaian ideas of the Earth as a self-regulating system, proposed by James Lovelock.

It was given expression through the CLAW hypothesis (published two decades ago by R. Charlson, J. Lovelock, M. Andreae & S. Warren), which supposes that the gas dimethyl sulphide produced by marine plankton influences cloud formation and hence albedo and climate.

However, direct evidence for a link between plankton and clouds has been slow to emerge. A recent paper (N. Meskhidze & A. Nenes *Science* **314,** 1419–1423; 2006) shows a tantalizing seasonal and spatial association between sea-surface chlorophyll (an indicator of biological activity) and atmospheric properties for a six-year period over a substantial area of the Southern Ocean.

Over high-chlorophyll areas, the number of cloud droplets doubled whereas the droplets' size decreased by 30% compared with other regions, leading to an atmospheric cooling comparable to that over highly polluted regions.

Meskhidze and Nenes attribute these changes to plankton emitting the gas isoprene. I am sceptical whether the sea-to-air flux of this compound is sufficient to produce the observed effects, but finding out what does give rise to the apparent association will keep me and other scientists involved in projects such as the Surface Ocean – Lower Atmosphere Study (www.solas-int.org) busy for many years.

It is vital to understand what is happening in order to be able to predict how future changes in biological activity in the oceans may mitigate or enhance climate change.

NEWS

Celebrity genomes alarm researchers

Genome researchers are questioning the plans of some of their number to stage high-profile releases of their very own genome sequences.

Tension over the issue surfaced this month at the annual genomics meeting at Cold Spring Harbor Laboratory in New York. There, some researchers expressed concerns that sequencing prominent scientists first will make personal genomics look like a tool for the rich and privileged.

At the meeting, Michael Egholm, a vice-president at 454 Life Sciences, a sequencing technology company in Branford, Connecticut, stood by a poster describing his company's effort to sequence the genome of genetics pioneer James Watson. The company claims this is the first sequence of an individual human genome, and that it took three months and cost about $1 million. "So, is this the next space tourism?" joked a scientist inspecting the poster.

Egholm winced at the implication that his company's plan is a vanity project for the wealthy. "What really matters is the *next* 100 genomes," he responded, carefully.

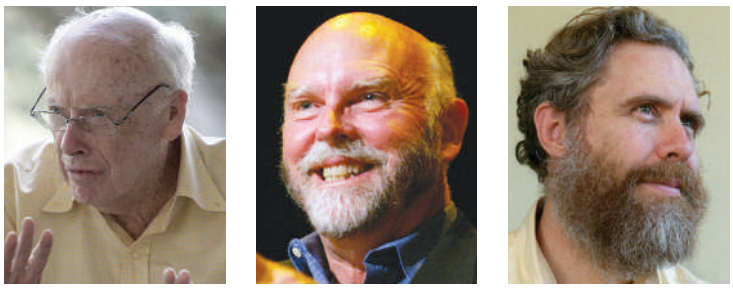
The exchange reflects tension between geneticists over whether sequencing scientists and celebrities is the best way to begin the long-promised era of personalized medicine. Watson and a handful of other famous people will be the first to see their complete genome sequences — and some scientists are uncomfortable with that.

"If all the sequences obtained over the next year or two are done on scientists with strong financial positions, that will send a message quite contrary to what the genome project aimed to achieve," says Francis Collins, head of the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland.

The sequencing of individual human genomes has long been a central goal of genomics. In 2001, the Human Genome Project produced a reference genome from the DNA of many different individuals. As with that project, attempts to sequence individual genomes are culminating in a race to finish first. On 31 May, 454 is expected to join scientists at Baylor College of Medicine in Houston, Texas, to present Watson with a copy of his genome.

But genomics pioneer Craig Venter claims he has already sequenced and assembled his own genome and submitted it to the publicly funded database GenBank. He says an analysis of his genome will be described in an upcoming paper in the journal *PLoS Biology*.

Next up will probably be sequencing guru George Church of Harvard University, who is one of the first ten volunteers for his privately funded Personal Genome Project. Then there is the Archon X Prize in Genomics — a $10-million cash award for the first team to sequence 100 genomes in 10 days — for which Venter is co-chair of the scientific advisory board. The prizewinner can claim a $1-million bonus by sequencing a list of 100 individuals, including people nominated by disease advocacy groups, and celebrities such as television journalist Larry King, cosmologist Stephen Hawking, Google co-founder Larry Page, Microsoft co-founder Paul Allen and former junk-bond trader Michael Milken.



The sequenced: (from left) James Watson, Craig Venter, George Church.

B. DUNYUAN/NEWSCOM/M. NAGLE/GETTY IMAGES/R. LINCOLN/HARVARD

The scramble has some researchers wondering whether the public will see personal genomics as an activity for the benefit of humanity. "I'd hate the availability of single-genome sequencing to be based purely on money and fame," says Michael Ashburner, a geneticist at the University of Cambridge, UK. "Just doing famous or very rich people is bloody tacky, actually."

"This is almost like recreational genomics, or the molecular equivalent of a whole-body scan, for those who have boundless curiosity and cash," says Kathy Hudson, director of Johns

Plans forge ahead for better weather monitoring

Meteorologists are planning a coordinated global drive to recalibrate space-based measurements of the weather. The weather scientists are confident that better calibration will result in better data — and a fuller picture of global climate change.

Meeting last week in Geneva, the World Meteorological Organization announced plans for a Global Space-based Inter-Calibration System (GSICS). The initiative will ask national satellite agencies to take steps to ensure better comparability of satellite measurements made by different instruments and satellites, and to tie these measurements to absolute references.

"As the requirement for monitoring global climate becomes clearer, there is need for more accurate measurements," says Don Hinsman, director of the World Meteorological Organization's space programme. "To permit early detection of climate change, it is vital that satellite instrument calibration is of the highest quality, and that a capability exists to cross-calibrate satellite sensors."

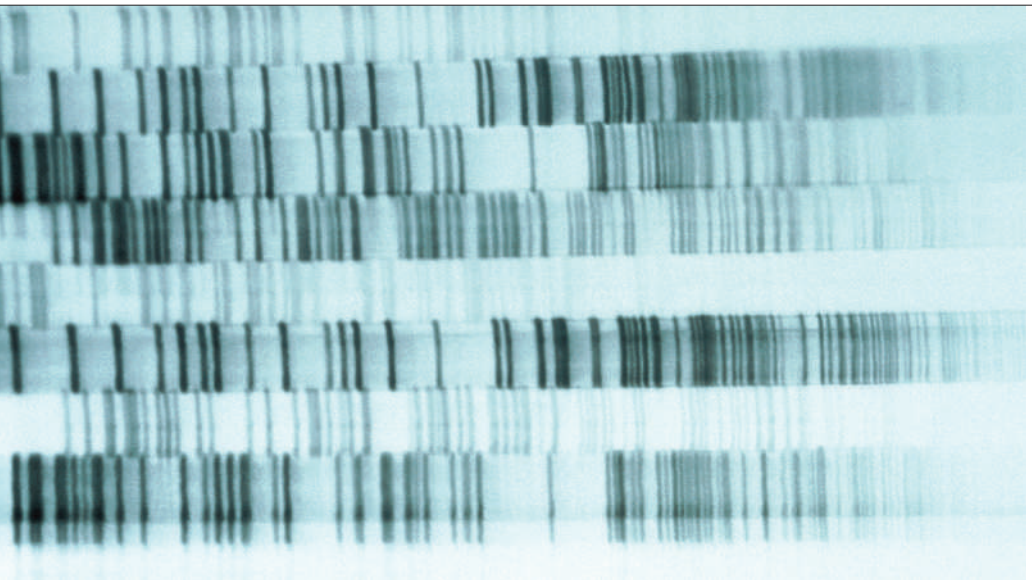
Remote sensing by some 30 satellites forms the backbone of global weather and climate monitoring today. Such measurements are vital because reliable ground-based observations are available for only about a quarter of Earth's surface. Continuous measurement of oceans, deserts and other remote and sparsely populated areas can come only from space.

But such measurements are prone to error, with problems arising from instrument degradation over time, small deviations of the satellites from their planned orbits, and faults in the algorithms used to process raw numerical data into meaningful geophysical information.

Flawed satellite data have caused disagreements between scientists in the past over such matters as temperature trends in the troposphere[1]. One radiometer onboard a US National Oceanic and Atmospheric Administration satellite — the only instrument to measure temperature in the stratosphere before 1998 — is thought to have transmitted grossly biased temperature measurements since 1979 (ref. 2).

NASA

HOW TO SURVIVE IN A BLACK HOLE
There's no escape, but how can you maximize your remaining time?
www.nature.com/news



J. WESTRICH/ZEFA/CORBIS

Some geneticists worry that sequencing the rich and famous is a misuse of genomics.

Hopkins University's Genetics and Public Policy Centre in Washington DC. "It will be sort of a sad statement if that's what we end up getting out of the Human Genome Project."

Scientists such as Venter counter that sequencing themselves first will defuse public anxiety about the possible misuse of genomic data by insurers, employers and others. The institutional review board that approved Church's project, for instance, said that only people with a master's degree in genetics, or the equivalent, should be allowed to volunteer, to ensure that they understand the implications.

Venter says he is creating a database where sequenced individuals can put genetic and personal information, such as medical records. He says he will deposit his own clinical information, possibly when his book *A Life Decoded: My Genome, My Life* is published in October.

"I think it would be wrong to take ten 'average' people and try to convince them it's all right for them to have their genetic code and life histories exposed on the Internet if people in a leadership position are not willing to do that," Venter says.

But it's not clear that all of the genome pioneers are acting altruistically. Watson said at the Cold Spring Harbor meeting on 10 May that he has not asked either of his grown sons for permission to publish his genome sequence, which 454 has said will be publicly posted in some form. That has raised questions about the responsibility of sequenced individuals to family members who share their DNA.

"This will be a challenging question, because if you're planning to put this information in a truly open database, there are issues of risk not just to you, but to your relatives," Collins says. "Jim clearly felt those risks were not such as to cause him to take action on them."

Watson knows that he and other individuals will not learn much of use from their own genomes — at least, not yet. Scientists are still learning to interpret genomic data, and have yet to unravel the genetic signatures of most diseases. There are some exceptions — Watson has asked 454 not to reveal the status of his gene for apolipoprotein E, because it is associated with Alzheimer's disease. But for the most part, personal genomes will not become useful until hundreds or thousands of individual genomes have been analysed.

The NHGRI is now planning to sequence about 100 individual genomes at its three publicly funded sequencing centres over the next couple of years. Collins says the institute will ask for scientific advice on who should be sequenced first. One question is what pool of sequenced individuals will yield the most useful information.

For instance, it might be possible to discover the basis of a rare genetic disease by sequencing many families affected by it. Scientists could also learn much from cancer patients or people who have already been studied in the International HapMap Project, a publicly funded effort to look at human genetic variation.

"We would want these to be chosen in such a way that you could get maximum information out of them, and there are various opinions about what that would mean," Collins says. ■

Erika Check

Even small temperature discrepancies, if undiscovered, can seriously disrupt the study of climate trends. "Inter-calibration has to be almost perfect if we want to look at climate trends — otherwise the bias will be stronger than the signal you want to address," says Jean-Noël Thépaut, who heads the satellite section at the European Centre for Medium-Range Weather Forecasts in Reading, UK.

The onboard calibration of instruments is costly and technically challenging, and provision for it has been incorporated only into new satellites. But just as important is the occasional lack of consistency between data collected from different satellite missions.

"The development of new sensor technology is progressing much faster than our capability to validate data," explains Gerhard Adrian,

"To permit early detection of climate change, it is vital that satellite instrument calibration is of the highest quality."

head of research at the German Weather Service in Wiesbaden.

The GSICS will make use of the exceptionally well-calibrated sensors onboard the latest generation of European and US meteorological satellites — such as Europe's MetOp-A satellite, which became operational last week — to validate data from older instruments.

"Cross-calibration is very much in our own interest," says Johannes Schmetz, head of the meteorology division at EUMETSAT, the European Organisation for the Exploitation of Meteorological Satellites in Darmstadt, Germany, and a member of a panel that will run the GSICS. "Ideally, what we would like to have is an operational system that could precisely define, and correct for, any orbital and instrumental biases in real time."

Reprocessing recently archived data using improved algorithms is also part of the plan. At EUMETSAT, robots can now do this quite quickly. Cumbersome manual 'data archaeology' is required only for old data sets stored on unwieldy magnetic tape.

Satellite data are becoming ever more abundant. At the European weather centre in Reading, for example, more than 5 million data points are processed every day, with the volume of data likely to triple in the next few years. ■

Quirin Schiermeier

1. Mears, C. A. & Wentz, F. J. *Science* **309,** 1548–1551 (2005).
2. Nash, J. & Edge, P. R. *Adv. Space Res.* **7,** 333–341 (1989).

Website homes in on climate hazards

Is the average US citizen ready to pay $30 to find out whether climate change will put their home at risk of being drowned by sea water, or burned in a wildfire?

David Purcell, a former banking executive, is betting that the answer is 'yes'. He has teamed up with climate scientists at the University of Arizona in Tucson and set up a website, called Climate Appraisal, that generates maps, graphs and commentary on environmental risks — including those related to climate change — for any location in the United States.

But some question whether, at this stage in the science, meaningful predictions on such a small scale can be made. "Ever since we knew about global warming, there's been a demand for bringing it down to a level that local people care about," says Gavin Schmidt, a climate modeller at NASA's Goddard Institute for Space Studies in New York, who hasn't seen Climate Appraisal's reports. "But there are some irreducible difficulties in doing so."

Purcell, who is based in Easton, Connecticut, says the website is a way to bring climate science to the public, and adds that his long-term goal is to direct at least 50% of the income it generates to climate research and education. He's built the site with the help of Jonathan Overpeck, a climate scientist at the University of Arizona and a lead author of the latest reports on climate change presented earlier this year by the Intergovernmental Panel on Climate Change (IPCC). "I saw that this could get people to take climate change seriously because they would see their stake in it," Overpeck says.

So far the undisclosed investment in Climate Appraisal has come from Purcell's pocket. Overpeck and two other Earth scientists at the University of Arizona have written articles for the site and advised on what data to present and how to present them for free. But Climate Appraisal has to strike a difficult balance in giving homeowners detailed information while not overstating the scientific certainty.

As the effects of global warming begin to be noticed, climate scientists have come under growing pressure to forecast what will happen on local scales of as little as a few kilometres. But most are wary of making statements about such impacts. It was only in its latest round of reports that the IPCC sought to do so — and it confined itself to producing forecasts for regions 500 kilometres across.

The Climate Appraisal website aims to forecast the likelihood of natural hazards.

S. MORTON/GETTY

To evaluate Climate Appraisal's approach, *Nature* asked six climate scientists to comment on five of the website's reports for a range of US locations.

Most said that the site's biggest plus was the way it collates so many types of information. Providing historical data on extreme events such as floods, wildfires, hurricanes and tornadoes, plus lists of toxic-waste sites, data on air quality, and information about certain diseases, is helpful to people who want a broad idea of an area's hazards, they say.

> "Ever since we knew about global warming, there's been a demand for bringing it down to a level that local people care about."

But the information that the site presents on climate change is limited, our reviewers said, in large part because Overpeck and his colleagues have been careful not to go beyond what the science tells them.

Climate Appraisal's main address-specific data on climate change are projected temperatures for 2050 and 2100 and, for coastal addresses, sea-level changes for the coming century and beyond. For example, users can zoom in on an area just a few kilometres across to see whether it is likely to be submerged if the sea rose by about a metre.

Philip Mote, one of the reviewers and a climatologist at the University of Washington in Seattle, says he would like to see more explanation and caveats given for both sea-level and temperature projections. In the case of sea-level rise, he says, although a metre rise is plausible within the next 100 years, complicating factors such as movement of Earth's continental plates make it difficult to predict how different shorelines will be affected.

"If I was doing this I would caveat it to death," says another reviewer, Linda Mearns, a climate scientist at the National Center for Atmospheric Research in Boulder, Colorado, and another lead author on the IPCC reports.

Overpeck recognizes that there are some gaps and glitches in the data, and is hoping that sales of the reports will allow him and his colleagues to replace some outmoded information with the latest IPCC forecasts and better models. For now, limited money and time mean using only the data that are "published, public and affordable". He says that the site's advice for users is as comprehensive as it needs to be, without being confusing. "If we included all the scientific discussion about all the uncertainties, no one would ever read it."

Purcell believes the site will give the public the kind of information that insurance companies are using to decide whether to cover certain risks. That's a live issue for many US homeowners: only last week, insurance company Allstate stopped offering policies to new

PLANE PILLS
How Viagra could beat jetlag
www.nature.com/news

homeowners in California because of the growing risk of wildfires there.

With decades of experience in risk assessment, insurance companies and the consulting firms that they use remain best-placed to evaluate a place's environmental hazards, says Celine Herweijer, principal climate scientist for Risk Management Solutions in London. She adds that although the idea of giving the public risk information is a good one, the information Climate Appraisal is currently selling will be of limited use to the homeowner.

In the case of floods, for example, homeowners would benefit most from estimates of the probability of a flood of a certain depth, Herweijer says, rather than from a national map of flood events, or a list of past floods in their area.

By linking climate change to someone's home address, Climate Appraisal is entering new and fairly treacherous waters. All in all, however, Herweijer and other climate scientists say the idea of making local risk information more widely available is valuable. In attempting to work out what to present and how to present it, most agree that the venture should be admired for its ambition. "You have to applaud it for what it is trying to do," says Schmidt. ■

Lucy Odling-Smee

Help flies in for human genome

Worms and flies are to be enlisted by researchers attempting to make sense of the instructions embedded in the human genome.

Since 2003, geneticists in the United States have been engaged in the pilot phase of a project called ENCODE — the Encyclopedia of DNA elements — which aims to catalogue all the functional parts of the human genome at a cost of around US$20 million a year.

Now, a four-year, $57 million project called modENCODE will add roundworms (*Caenorhabditis elegans*) and fruitflies (*Drosophila melanogaster*) to the mix. By analysing these trusty model organisms, scientists hope to make better sense of the human-genome data gleaned from ENCODE.

Grants from modENCODE will support scientists who are cataloguing the major types of functional genetic elements. These include RNAs that cut genetic transcripts into different genes and fine-tune gene expression, modifications to bundles of protein and DNA called chromatin, and DNA sequences that control how and when genes are transcribed.

Researchers say that the project, which was announced by the US National Human Genome Research Institute (NHGRI) on 14 May, is needed because ENCODE has shown that the human genome is even more complex than they originally thought (see *Nature* **441,** 398–401; 2006).

ENCODE's pilot phase covered just 1% of the human genome and has generated a slew of techniques for analysing the genome. "We now have a portfolio of technologies that can be applied with high specificity and sensitivity to decode the parts list of the human genome at a rather precise level," says Francis Collins, the NHGRI's director.

So far, the project has revealed that genes are regulated by complicated networks that span huge portions of the genome; that RNA plays a previously unappreciated part in determining how proteins are made from genes; and that scientists don't fully understand some of the regulatory motifs that they have been studying for years.

Scientists hope that modENCODE will enable them to work out these processes by looking at the model organisms, which have much smaller genomes than humans and are easier to manipulate in the lab. The roundworm has about 100 million base pairs in its genome and the fruitfly 180 million, compared with 3 billion base pairs found in the human genome.

"These are hard issues, and without the experimental feedback you can get from worms and flies, it's hard to see when we're really going to touch ground," says Robert Waterston of the University of Washington in Seattle and recipient of a $5.4-million modENCODE grant to study *C. elegans*.

But at the end of modENCODE's first four years, will scientists be able to pat themselves on the back and move on to something else? It's unlikely, they say, given that human genetics seems to evolve into a more complicated subject with every new foray into the genome.

"We should get a lot closer," says Lincoln Stein of Cold Spring Harbor Laboratory in New York, and the head of a new centre to coordinate the modENCODE data. "If we can't nail it down, we'll at least have a lot of it thumbtacked to the wall." ■

Erika Check

K. TAYLOR/NATUREPL.COM

Analysing the fruitfly could help researchers uncover the secrets of the human genome.

NUMBER CRUNCH

89,826 people attended the FA Cup Final on 19 May, the showcase event of the English football season.

688 food and drink outlets served hot dogs and beer to the hordes of fans at the brand new Wembley Stadium.

3,000 football pitches would fit into the ecological footprint of the event, according to researchers at Cardiff University — an estimate of the land area needed to make the food and drink, and to generate the energy for getting everyone and everything to the game.

ZOO NEWS

Soft shell

Conservationists have found "an abundance" of Cantor's giant soft-shelled turtles (*Pelochelys cantorii*) at a former Khmer Rouge stronghold on the Mekong River in Cambodia, calming fears that the creature is nearly extinct.

CONSERVATION INTERNATIONAL/AP

SCORECARD

▲ **Office workers**

If your deskbound existence is making your waistline spread, never fear — the inventors of the 'office treadmill', which lets you work and walk at the same time, promise that users could lose up to 30 kilograms in a year.

▼ **Estonian businesses**

The small Baltic country seems to be the victim of the world's first large-scale act of cyber war, as almost all government computers and many banking facilities were recently paralysed for several weeks by hackers. The highly coordinated nature of the disruption, coupled with recent tensions with Russian expatriates in Estonia, has led some to say that Russia was behind the attack.

Sources: The Guardian, BBC, ABC News

SIDELINES

Japan centres aim to put science in premier league

Japan's top researchers are this week putting the finishing touches to their applications to run a World Premier International Research Center, the grandiose title of the latest government effort to boost Japanese scientists' links with their colleagues overseas.

There will be some five centres, each of which will receive base funding of between US$4 million and $16 million for up to ten years. They are the latest in a long line of efforts to make Japanese research more flexible and international in outlook.

Isolation is a major factor weakening Japanese science, says neuroscientist Takao Hensch, who last year moved to Harvard University after ten years at the RIKEN Brain Science Institute in Wako, north of Tokyo. "To be competitive there must be constant communication ensuring that Japanese scientists are respected participants in their fields," Hensch says.

"The important thing is being equal. It's hard to change these things with compulsory rules."

Japanese policy-makers, including officials at the science ministry, accept that the country's science is isolated as a result of its culture, geography and language. They also suspect that Japanese science is underperforming as a result.

The new institutes, which will be selected in September by an international review panel, are expected to serve as 'globally visible' research centres and to attract top-level researchers from around the world. To prevent the centres from merely paying lip-service to these goals, the application stipulates that 30% of the roughly 200 researchers expected at each centre and 10–20% of the 10–20 principal investigators must be foreigners. Lack of progress towards these goals could lead to closure.

Such attempts to make Japanese research more international are not new. The Okinawa Institute of Science and Technology, for example, which officially opened this year, had said that it would hire half of its research staff from abroad and looks set to meet that target. At the RIKEN Center for Developmental Biology (CDB) in Kobe, which opened in 2002, 10% of the staff and principal investigators are foreigners. And 20% of the staff at the RIKEN Brain Science Institute, which was set up in 1997, are also from overseas.

But these institutes have remained exceptions in a conservative Japanese system that is generally not regarded as being welcoming to non-Japanese scientists.

Some of the problems lie outside the institutes themselves. Douglas Sipp, who heads the CDB's international-relations office, says that although all research material is available bilingually, the ministries tend to send documents such as grant-programme notices in Japanese, with English versions sometimes arriving too late to be of use. Sipp also says that the difficulty foreign researchers have in paying for international schools for their children or

NIH presents the mind of a child

A US National Institutes of Health (NIH) study into brain development during childhood is creating a database as a benchmark against which researchers on other studies can compare their data.

Scientists will be able to apply to the NIH for access to the database once it is ready, probably before the end of the year. They could, for example, compare structural or behavioural data from their studies of a mental disorder against those for children with 'normal' brains.

The study, called the MRI Study of Normal Brain Development, is building a database of what constitutes a 'normal' child. Five hundred children aged from 7 days to 18 years, and representative of US society, have been recruited from six centres.

SOVEREIGN, ISM/SPL

A database of scans will offer a picture of a normal child's brain.

Those with, or at risk of, any neurological or psychiatric

M. HENLEY/PANOS

Despite massive investment, Japan still lags behind other developed nations in scientific productivity.

finding posts for their spouses makes recruitment hard.

Hensch says the situation in the universities remains a major stumbling-block. "Although it is fine to provide great resources, autonomy and accountability for talented young researchers, without similar openings in the traditional hierarchical university system there can be no culture of mobility and turnover."

Tasuku Honjo, an immunologist at Kyoto University, agrees that the universities must change. Honjo is a member of the country's highest scientific decision-making body, the Council for Science and Technology Policy, which introduced the latest initiative. Reform of the universities in 2004 (see *Nature* **419,** 875–876; 2002) was meant to give them the freedom to compete both with each other and internationally for top talent. "In actuality, nothing has changed," Honjo says.

The problem, Honjo adds, is that Japanese institutions continue to place fairness above excellence. "The important thing is being equal," he says. "It's hard to change these things with compulsory rules." But the new centres are intended to become models for how freedom should be exercised.

Officials hope the centres will play a major role in boosting Japan's scientific productivity. Figures released by Thomson Scientific on 15 May ranked Japan second in the world for the number of scientific publications between 1996 and 2006. But Germany, which over that period invested half the funds in science compared with Japan, had only 6% fewer papers. And for the number of high-impact papers, Japan ranked only fifth.

The interpretation of such data is, of course, open to debate. Some observers point out, for example, that Japan's research investment may be bearing more fruit through its highly successful and innovative industrial corporations than through its number of publications. ■

David Cyranoski

disorders were screened out of the study. All were given a magnetic resonance imaging (MRI) scan, which shows the size of the brain's structures and the densities of grey and white matter, as well as tracks of fibres connecting areas in the brain. The children also took behavioural and cognitive tests.

The first results of the study, on neuropsychological tests in 6–18-year-olds, were published on 18 May (D. P. Waber *et al. J. Int. Neuropsychol. Soc.* doi:10.1017/S1355617707070841; 2007). They show that cognitive skills improve between the ages of 6 and 10 but level off during adolescence — contradicting a widespread belief that cognitive development 'spurts' during adolescence.

Researchers also confirmed that children's abilities in some cognitive tasks differ between boys and girls, and that cognitive performance correlates positively with parental income. But the differences were smaller than those seen in other studies. "We don't know why," says NIH project officer Katrina Gwinn. "It may be the way we selected our sample, or we may know less about biases in our psychological tests than we like to think."

More data will be added when later parts of the study are analysed. First data from another part of the study, involving around 100 babies aged up to 4.5 years scanned at more frequent intervals, will probably be published before the end of the year. ■

Alison Abbott

Darwin sceptic says views cost tenure

He's a young astronomer with dozens of articles in top journals; he has made an important discovery in the field of extrasolar planets; and he is a proponent of intelligent design, the idea that an intelligent force has shaped the Universe. It's that last fact that Guillermo Gonzalez thinks has cost him his tenure at Iowa State University.

Gonzalez, who has been at Iowa State in Ames since 2001, was denied tenure on 9 March. He is now appealing the decision on the grounds that his religious belief, not the quality of his science, was the basis for turning down his application. "I'm concerned my views on intelligent design were a factor," he says.

Advocates of intelligent design are rallying behind Gonzalez in the latest example of what they say is blatant academic discrimination. "Academia seems to be in a rage about anything that points to any purpose," says Michael Behe, a biochemist and prominent advocate of intelligent design at Lehigh University in Bethlehem, Pennsylvania. "They are penalizing an associate professor who's doing his job because he has views they disagree with."

But other researchers think that the department's decision was entirely justified. "I would have voted to deny him tenure," says Robert Park, a physicist at the University of Maryland in College Park. "He has established that he does not understand the scientific process."

Gonzalez's early career was far from controversial. He graduated with a PhD from the University of Washington, Seattle, in 1993 and did a postdoc at the University of Texas in Austin. "He proved himself very quickly," says David Lambert, director of the university's MacDonald Observatory. He and Gonzalez co-authored several papers on variable stars, and Lambert says that while there, the young Cuban immigrant was an impressive scientist. "He is one of the best postdocs I have had," he says.

"Gonzalez is one of the best postdocs I have had."

In 1996, Gonzalez returned to the University of Washington to do his second postdoc, and again distinguished himself — producing two papers[1,2] that linked a star's metal content to the presence of extrasolar planets around it. The papers are still highly cited, and they have encouraged other researchers to search for planets around metal-rich stars.

The 43-year-old astronomer is also a deeply religious evangelical Christian, and his faith has shaped his views on science. He considers himself a "sceptic" of Darwin, and says that his Christianity helps him to understand Earth's position in the Universe. "Our location in the Galaxy, which is optimized for habitability, is also the best place for doing cosmology and stellar astrophysics in the Galaxy," he says. In other words: "The Universe is designed for scientific discovery."

N. MAJUMDAR/AP

Astronomer Guillermo Gonzalez has been actively promoting his book on intelligent design.

Gonzalez refrained from mentioning his beliefs in his teaching and peer-reviewed works, but in 2004, he co-authored a book entitled *The Privileged Planet*, which included many of his pro-design arguments[3]. He has since travelled the country delivering talks that support the thesis of his book.

His work did not go unnoticed at Iowa State. In 2005, Gonzalez's rising profile led a group of 131 faculty members to sign a petition disavowing intelligent design. "We were starting to see Iowa State mentioned as a place where intelligent-design research was happening," says Hector Avalos, a religious-studies professor who helped lead the signature drive. "We wanted to make sure that people knew the university does not support intelligent design." Avalos adds that they did not name Gonzalez directly, and he takes no position on the astronomer's tenure.

Nevertheless, proponents of intelligent design point to the signature drive as evidence of a widespread academic hostility to those who support the idea. "There is a pattern happening to everybody who's pro intelligent design," says one pro-design biologist, who declined to be named because his own tenure process has just begun. "The same thing could happen to me," he says. "I don't want to get into trouble."

But Park says that a researcher's views on intelligent design cannot be divorced from the tenure decision. Anyone who believes that an intelligent force set the Earth's location doesn't understand probability's role in the Universe, Park argues. Such a person is hardly qualified to teach others about the scientific method. "We're entrusting the minds of our students to this person," he says.

But not all scientists agree. "Nothing I have seen in his refereed papers leads me to believe his beliefs are impinging on his science," says David Lambert. "I would have said he was a serious tenure candidate."

Eli Rosenberg, who chairs Iowa State's physics department, concedes that Gonzalez's belief in intelligent design did come up during the tenure process. "I'd be a fool if I said it was not [discussed]," he says. But, he adds, "intelligent design was not a major or even a big factor in this decision." Four of twelve tenure candidates have been turned down in the past decade, he says. "We are a fairly hard-nosed department."

Iowa State's president Gregory Geoffroy is now reviewing Gonzalez's appeal. He has until 6 June to make his final decision. ■

Geoff Brumfiel

1. Gonzalez, G. *Mon. Not. R. Astron. Soc.* **285,** 403–412 (1997).
2. Gonzalez, G. *Astron. Astrophys.* **334,** 221–238 (1998).
3. Back, C. H. & Pescia, D. *Nature* **428,** 808–809 (2004).

Concern over teen suicides extends flu-drug probe

Japan is widening its investigation into whether certain influenza drugs could have dangerous side effects, including psychiatric problems and suicidal tendencies, in certain groups of people.

In March, the Japanese health ministry advised doctors that they should not prescribe teenagers the flu drug Tamiflu (oseltamivir) — made by Roche — after reports of some young people on the drug throwing themselves from buildings (see *Nature* 446, 358–359; 2007). Last week, the health ministry said that it would also look into the flu drugs Relenza (zanamivir) — made by GlaxoSmithKline — and amantadine.

A study that ended in 2006 flagged up all three drugs for their potential side effects, but found no clear evidence that Tamiflu was to blame for the teenagers' abnormal behaviour. The latest investigation is expected to report its results by this winter's flu season.

Power cut endangers marine lab experiments

An island research lab off the coast of California scrambled to save a number of marine experiments earlier this month because it had no back-up generators when an unexpected wildfire shut down power.

The Wrigley Marine Science Center, on Catalina Island about 30 kilometres from Los Angeles, lost power shortly after the 10 May brush fire burned down power poles. Operated by the University of Southern California, the centre's biggest loss was about 100 million oyster larvae, which were being used in a genomics study to examine gene expression in various environments. With power out, electrical seawater pumps could not be operated, so the larvae were put to sea.

Staff members imported boatloads of dry ice to save a decade's worth of frozen specimens.

US gives green light to rice with breast-milk proteins

The US Department of Agriculture (USDA) has approved one of the first large-scale plantings of a food crop genetically modified to contain human proteins. The crop will be planted in Kansas.

Ventria Bioscience in Sacramento, California, has made strains of rice that produce proteins found in breast milk — lysozyme, lactoferrin and human serum albumin. Ventria says it aims to use the rice to make drinks that can combat diarrhoea, and dietary supplements to treat anaemia.

The USDA approved the crop after receiving 29 positive comments from the public and 20,005 negative ones. In the end, the department decided that fears the rice would escape into the environment or the food supply were not warranted.

Telescope on a plane takes to the skies — at last

After years of being grounded by budgetary delays, the US/German Stratospheric Observatory for Infrared Astronomy (SOFIA) is getting airborne at last.

The 2.5-metre telescope is mounted aboard a Boeing 747 that will fly above atmospheric water vapour. NASA rededicated the plane, the *Clipper Lindbergh*, on 21 May, the eightieth anniversary of Charles Lindbergh's solo flight across the Atlantic.

C. THOMAS/NASA

Airborne observatory SOFIA is several years late.

In the past month SOFIA has successfully completed its first two test flights, says Eric Becklin, chief scientist for the project at the Universities Space Research Association, the Columbia, Maryland-based nonprofit organization overseeing the project.

SOFIA, which cost US$700 million, will soon move from Waco, Texas, where it was built, to NASA's Dryden Flight Research Center at Edwards Air Force Base in California, for further tests. It is expected to make its first science observations in early 2009.

Creationist museum to open in Kentucky

American families looking for an educational activity this summer could find themselves lured into the newest and splashiest museum in Petersburg, Kentucky — a $27 million Creation Museum.

Minister Ken Ham and his Answers in Genesis group say they built the museum as a counterpoint to traditional science and natural-history museums. Exhibits include baby *Tyrannosaurus* playing with human children, and a reproduction of part of the Grand Canyon, explained as being formed by Noah's flood.

There is a Museum of Creation and Earth History in Southern California, but the Kentucky museum is larger and expects to draw families on themed outings.

Protesters are planning a 'Rally for Reason' at the museum's 28 May opening. Science advocates have complained, saying the museum misrepresents scientific understanding of Earth history.

Virtual journey to the centre of the Earth

The Earth Simulator Center in Yokohama, Japan, has a new attraction along with its supercomputer: a mini-theatre that will allow scientists to interact with data in three dimensions.

The theatre extends the experience of working with complex dynamic systems. "You can place a massless virtual dust particle in the middle of a typhoon and watch what happens to it in the same way that a golf player drops grass to test the wind," says Akira Kageyama, a simulation specialist at the centre. He hopes the facility will open to all Earth Simulator users in the near future.

Visitors have been known to duck to avoid oncoming objects thrown by a virtual typhoon or even to run into the walls while viewing the inside of Earth's core.

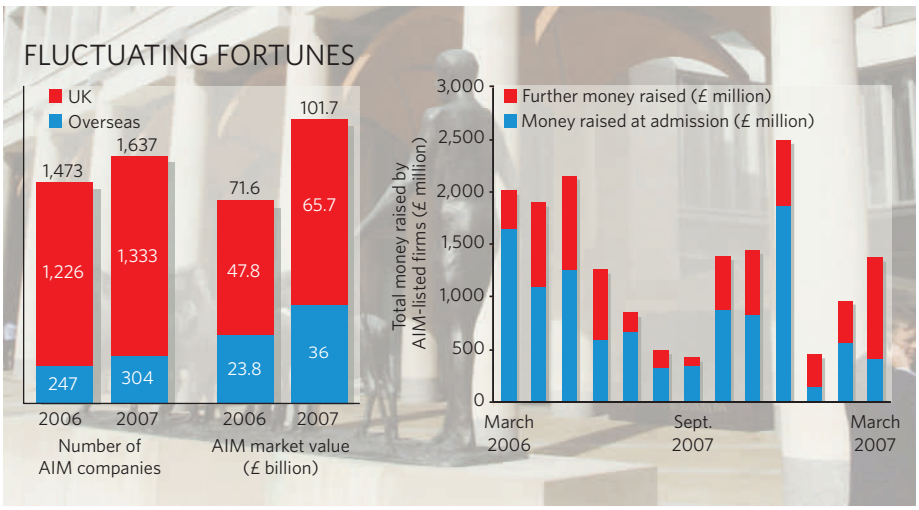
Correction

The News story 'Wind farms' deadly reputation hard to shift' (*Nature* **447,** 126; 2007) said that on average each wind turbine kills 0.03 birds every year. In fact, this is the figure only for birds of prey. The overall average for all birds is 4.27 deaths per turbine per year.

BUSINESS

London calling

Despite its critics, the Alternative Investment Market could still be attractive to America's small, innovative companies, reports **Andrea Chipman**.



F. HANSON/PA

Compared with Nasdaq, the main American exchange for high-growth companies in biotechnology and similar fields, the Alternative Investment Market (AIM) of the London Stock Exchange is a bijou affair. In March, the most recent month for which data are available, the 1,637 companies listed on AIM had a combined market value of £102 billion (US$201 billion, see graph). Nasdaq, in its April report, boasted twice as many companies and almost 20 times the market capitalization: $3.8 trillion.

Yet last year AIM brought off some impressive coups. A string of American high-tech companies, including Aqua Bounty Technologies of Waltham, Massachusetts, which produces hybrid fish and vaccines for fish farms, and Entelos of Foster City, California, which concentrates on *in silico* disease models and 'virtual patients', chose the London market for their initial public offerings. It seemed that this run might continue as more small American companies, which are searching farther afield for capital, turned to the London exchange. As yet, though, there has been little activity of this sort, and AIM has come in for some serious criticism from US financial grandees.

In January, New York Stock Exchange chief executive John Thain told the World Economic Forum in Davos, Switzerland, that AIM's lax regulatory approach threatened the reputation of London as a capital market. Soon after, Roel Campos, a commissioner at the US Securities and Exchange Commission, was reported by Dow Jones Newswires as likening AIM to a casino, and billionaire investor Wilbur Ross weighed in on 18 April, telling the *Financial Times* that the AIM market was "clearly a dangerous one".

AIM's defenders attribute the attacks to the financial rivalry between New York and London. A prominent issue is the contrast between AIM's listing requirements and the more stringent ones imposed on US markets by the Sarbanes-Oxley (SOX) corporate governance legislation. SOX, which requires companies to implement detailed internal accounting and auditing structures, has made US public offerings more complex and expensive; companies say it can cost up to $3 million just to put the regime in place, and around $1 million annually to remain compliant.

> **"The barriers to US public markets for biotech companies are extreme."**
> — Scott Morrison

SOX is not the only issue. "The barriers to US public markets for biotech companies are extreme," says Scott Morrison, US biotechnology practice leader at consultants Ernst & Young. He adds that many institutional investors in America are limiting their sights to "companies that have checked all the boxes" — those already boasting an impressive list of investors, with a product in later-stage clinical trials and, increasingly, already in partnership with a larger biotech or pharmaceutical company. These thresholds are hard for early-stage life-science companies to cross.

Corporate governance complaints about AIM are aimed fairly specifically at the listing of companies from some emerging markets rather than US biotech firms. But AIM has other drawbacks that are more germane to those firms. Critics point to a shortage of market specialists and investors familiar with the life-science sector in London. AIM seemed to acknowledge such worries in February, when it issued a new rulebook for its 'nominated advisers,' or Nomads: financial institutions that shepherd companies through the flotation process. The new rules specified that Nomads must have "adequate knowledge of the business and sectors in which their AIM companies act", adding that "this will particularly be in relation to technical or complex areas such as biotechnology, but could also extend to a wider range of sectors".

There are other disincentives, too. Shares listed on AIM are not as easily transferable to US investors as shares on Nasdaq, making it harder for investors in American AIM-listed companies to sell their shares. The amount of money companies can raise on AIM is generally less than would be available on US exchanges, so some companies might be able to raise more by staying privately held and appealing to venture capital. Moreover, the success of foreign companies on AIM in tapping the market for a second or third time remains mixed. With US companies no longer a novelty, some investors may have had their fill, says Bruce Jenett, a California-based partner in corporate finance at international law firm Heller Ehrman.

Despite these possible drawbacks, many smaller US companies still look positively at AIM and other foreign markets, according to Matt Gardner, president of BayBio, a biotech trade group in northern California. Listing on AIM offers them not just a way of raising money, but also of enhancing the company's profile; many AIM companies could be positioned for an eventual Nasdaq listing, company advisers and investors say.

The small size of the market is still an issue, though. It would only take problems with a few of the listed companies to lead to trouble. "There was this massive gold rush a year ago, a huge influx of companies whose fundamental business is very complicated," says Jennett. "If one of them stumbles, the entire sector may be dragged down and frozen in terms of anyone wanting to put money into it." ■

The molecular wake-up call

It is 50 years since Arvid Carlsson showed dopamine to be a neurotransmitter. **Alison Abbott** profiles a chemical and its champion.

They were conscious but you wouldn't know it: able to perceive the world around them but powerless to look around, sniff the air or to cry out. So when the young scientist injected them with a chemical called L-dopa, he witnessed what seemed to be a miracle. They stirred, opened their eyes and began roaming around as if nothing had happened.

This may sound familiar from the book *Awakenings*[1] — the true story of how, in 1963, the neurologist Oliver Sacks used L-dopa to spectacularly revive patients with sleeping sickness who had been 'frozen', speechless and motionless, for more than 40 years. But the unwritten and equally startling prequel took place in Lund, Sweden, several years earlier. The protagonists were rabbits; their saviour a young Swedish pharmacologist called Arvid Carlsson.

In his experiment, Carlsson showed that dopamine — the chemical manufactured from levodopa, or L-dopa — acts as a neurotransmitter in the brain, passing signals between neighbouring neurons. Injection of L-dopa restored the propagation of electrical signals in the brains of rabbits that had been rendered catatonic, allowing the animals to move. But the pharmacological establishment was scornful of Carlsson's claim. At a London meeting in 1960, the foremost experts in neural transmission made it clear that they didn't believe him — dopamine was thought to be the metabolite of another neurotransmitter rather than one in its own right.

Within years the critics were silenced. Dopamine was shown to be a pivotal chemical in the neural circuits that drive pleasure and addiction, as well as in illnesses such as Parkinson's disease, for which L-dopa quickly became a first-line treatment. It remains so today. In 2000, Carlsson shared the Nobel Prize in Physiology and Medicine for his discovery. And next week neuroscientists will gather at a meeting in Carlsson's home town of Gothenburg, Sweden, to celebrate the 50th anniversary of his formative paper on the awakened rabbits[2].


1200

NATURE November 30, 1957 VOL. 180

ploid elements is homogeneous population as to their deoxyribonucleic acid content.
This work was aided by a grant of the Belgium F.N.R.S.

R. LAQUERRIÈRE

Belgian Centre of Growth and Differentiation, Department of Human and Comparative Anatomy, University of Ghent.

[1] Gerneli, G., *Arch. Biol.*, **68**, 1 (1957).
[2] Laquerrière, R., *C.R. Soc. Biol.* (in the press).
[3] Lison, L., *Acta Anat.*, **10**, 333 (1950).

3,4-Dihydroxyphenylalanine and 5-Hydroxytryptophan as Reserpine Antagonists

THE depletion by reserpine of storage in the body of 5-hydroxytryptamine ('Serotonin') and of the catechol amines is now well established[1–3]. In reserpinized animals the peripheral part of the adrenergic system does not function owing to lack of the transmitter[4]. This is presumably true also of the central part of the adrenergic system. However, it remains to be proved to what extent the central action of reserpine may be attributed to changes in brain catechol amines and/or 5-hydroxytryptamine.

If lack of amines were responsible for the central action of reserpine, administration of the amines in question should counteract these effects, provided that the amines were capable of entering the brain. However, 5-hydroxytryptamine has been shown not to penetrate the blood-brain barrier readily[4], and this may be true also of the catechol amines. This difficulty may be overcome by administering the amino-acid precursors of the amines. Thus injection of 5-hydroxytryptophan is followed by an increase in the level of 5-hydroxytryptamine in brain as well as by central excitation[4]. Preliminary experiments in this laboratory have shown that in this respect 3,4-dihydroxyphenylalanine, which is the precursor of the catechol amines (dopamine, noradrenaline, and adrenaline), behaves similarly.

Experiments were performed on mice (males weighing about 10 gm.), which received an intraperitoneal injection of reserpine (20–40 mgm. per kgm.). After about 16 hr., when the animals were markedly tranquillized and showed complete ptosis of the eyelid, 5-hydroxytryptophan, 3,4-dihydroxyphenylalanine, or a mixture of both (in the DL-form throughout) [illegible] tonically. In [illegible] 5-hyd[illegible]

A dramatic effect of 3,4-dihydroxyphenylalanine (200 mgm. per kgm. intravenously) was observed also in rabbits which had received reserpine in a dose of 5 mgm. per kgm. intravenously 4 hr. earlier. Within 10–15 min. after the injection of 3,4-dihydroxyphenylalanine the tranquillization as well as ptosis and miosis caused by reserpine had disappeared completely. If the animal had received iproniazid (100 mgm. per kgm. intravenously) about 2 hr. before the 3,4-dihydroxyphenylalanine, the dose of the latter required to antagonize the effect of reserpine was markedly reduced. This supports the assumption that the effect of 3,4-dihydroxyphenylalanine was due to an amine formed from it. (The iproniazid, when administered about two hours after the reserpine as in those experiments, did not *per se* counteract the tranquillizing effect of reserpine.) In normal rabbits 3,4-dihydroxyphenylalanine caused central stimulation, which was likewise markedly potentiated by iproniazid pretreatment.

A full account of these experiments will be published elsewhere.

ARVID CARLSSON
MARGIT LINDQVIST
TOR MAGNUSSON

Department of Pharmacology, University of Lund, Lund.
June 19.

[1] Shore, P. A., Pletscher, A., Tomich, E. G., Carlsson, A., Kuntzman, R., and Brodie, B. B., *Ann. N.Y. Acad. Sci.*, **66**, 609 (1957).
[2] Carlsson, A., Rosengren, E., Bertler, Å., and Nilsson, J., Internat. Symp. on Psychotropic Drugs, May 9–11, 1957 (in the press).
[3] Shore, P. A., and Brodie, B. B., Internat. Symp. on Psychotropic Drugs, May 9–11, 1957 (in the press).
[4] Udenfriend, S., Weissbach, H., and Bogdanski, D. F., *Ann. N.Y. Acad. Sci.*, **66**, 602 (1957).

Antihypertensive Activity of Hexahydro-1-Azepinepropionamidoxime

HEXAHYDRO-1-AZEPINEPROPIONAMIDOXIME dihydrochloride (I), designated as *SU*-4029, has been studied for its effect on the cardiovascular system of the dog.

(I)

[illegible] single intravenous dose of 30 mgm./kgm. of this [illegible] and lowered the arterial pressure of neurogenic [illegible] hypertensive dogs while not notably [illegible] the blood pressure of normotensive dogs. [illegible], in normotensive animals 30 mgm./kgm. of [illegible] given intravenously eliminated the severe [illegible] elicited by high doses of amphetamine [illegible] and also markedly antagonized carotid [illegible] reflex pressor responses. These antihyper- [illegible] effects were slow in onset and lasted for [illegible] daily two to six weeks following single [illegible] *SU*-4029 was orally active and had a [illegible] action when given in small daily doses. [illegible] the compound, hexahydroazepine was [illegible] ded and gave hexahydro-1-azepinepro- [illegible] b.p. 121–123°/14 mm.; n_D^{20} 1·4710 [illegible] 20·82; H, 10·52; N, 18·53 per cent. [illegible] res: C, 71·11; H, 10·61; N, 18·43 [illegible] eatment of this nitrile with hydroxyl-

Catatonic rabbits were revived by dopamine in a 1957 experiment led by Arvid Carlsson.

During the past half century, Carlsson and dopamine have followed intertwined paths. Researchers now understand that the way dopamine works is subtle and complex, and its mechanisms of action are central to the function of many neurological and psychiatric drugs. And Carlsson, now a sprightly 84-year-old, still spends hours pondering the mysteries of brain chemistry. But he feels marginalized in Gothenburg and, last year, the institute established in his name closed prematurely after bitter feuds about funding.

The field of biomedicine has also evolved during this time, and much has changed. "We used slide rules and manual calculators back then, so statistical calculations were quite time consuming," Carlsson says. But a willingness to embrace new techniques and fight for controversial ideas, as Carlsson has done throughout his career, remains essential to success. "He was one of my heroes when I was a young neurology resident," says William Langston, head of the Parkinson's Institute in Sunnyvale, California. "What impressed me was his brilliant deductive reasoning."

Carlsson is a polite, soft-spoken and humorous gentleman with a hallmark bow tie. He grew up in a family of academics in Lund that focused on the humanities but, in what he refers to as a minor act of rebellion, he chose science over the arts because he thought it more useful to society. He is also single-minded and uncompromising, features that no doubt contributed to his success as a scientist against stiff opposition — and perhaps also to his later problems.

Carlsson calls himself a lucky man, and says his first stroke of luck came when he arrived in the lab of chemist Bernard 'Steve' Brodie at the National Institutes of Health in Bethesda, Maryland, in the summer of 1955. Brodie's lab

pioneered studies with reserpine, one of the very first drugs to be introduced for the selective treatment of schizophrenia, and hence one of the hottest molecules for pharmacologists of the day. Reserpine injections made rabbits catatonic, but how the drug worked was a mystery.

Direct measures

At the time, pharmacologists typically tested the potency of neurotransmitters with assays of their biological activity — for example, applying them to a piece of animal gut under tension in an organ bath to see how much they could contract or relax the muscle. Brodie's team instead developed the spectrophotofluorimeter, a machine able to measure how much neurotransmitter was synthesized from fluorescently tagged precursors. This allowed the researchers to measure the precise level of a compound extracted from tissue rather than an indirect measure of its activity, and later became a standard instrument in biological labs. When Carlsson returned to Lund after his five-month visit to Brodie's lab, the first thing he did was order a spectrophotofluorimeter. "I didn't want to be confined by the indirectness of bioassays," he says.

Carlsson's work with this device showed that reserpine completely drains the stores of several neurotransmitters in the brain. Loss of one of these was causing the rabbits to become catatonic — the question was, which one?

Carlsson reasoned that he could answer this question by adding back the missing neurotransmitter to rabbits that had been frozen with reserpine — the crucial awakening experiment. The blood–brain barrier prevents the neurotransmitters noradrenaline and serotonin from passing into the brain from the blood, so Carlsson instead injected precursors of these molecules that can enter the brain and are then metabolized into the relevant neurotransmitter. One of these precursors was L-dopa, which is converted into dopamine and then, in turn, into noradrenaline.

But when Carlsson examined the revived rabbits' brains after injecting L-dopa he saw a lot of dopamine and very little noradrenaline. At this point it dawned on him that dopamine could be a neurotransmitter in its own right, a memory that still summons astonishment to his face.

Within months his graduate students Åke Bertler and Evald Rosengren had found that dopamine is normally concentrated in areas of the brain known to be involved in movement, such as the basal ganglia[3]. Knowing that high doses of reserpine cause side effects that are similar to some of the movement difficulties experienced by patients with Parkinson's disease, Carlsson proposed that this disease might be caused by a lack of dopamine.

"I won the Nobel Prize 40 years after my discovery. Einstein won one some 20 years after his. So I guess my work was twice as complicated."
— Arvid Carlsson

Sparking opposition

In late 1958, Carlsson travelled across the Atlantic to explain his ideas to a symposium in Bethesda. There, his story was well-received. "But this was not the case when I presented my results at the Ciba meeting in London," says Carlsson, who is still clearly a bit hurt.

The prestigious 1960 Ciba Symposium on Adrenergic Mechanism attracted all the key European players in the field. At the time, a vigorous debate was going on between the 'soups' — who thought that nerve transmission occurred through chemicals — and the 'sparks', who argued that it was all electrical. The soups had more or less won their case for neurotransmission outside the brain but, owing to lack of experimental evidence, the sparks' view still dominated the brain itself. "It's hard to imagine now, but when I was an undergraduate student at Cambridge [in the late 1950s] we were taught categorically that there was no chemical transmission in the brain — that it was just an electrical machine," recalls pharmacologist Leslie Iversen, professor emeritus at the University of Oxford, UK.

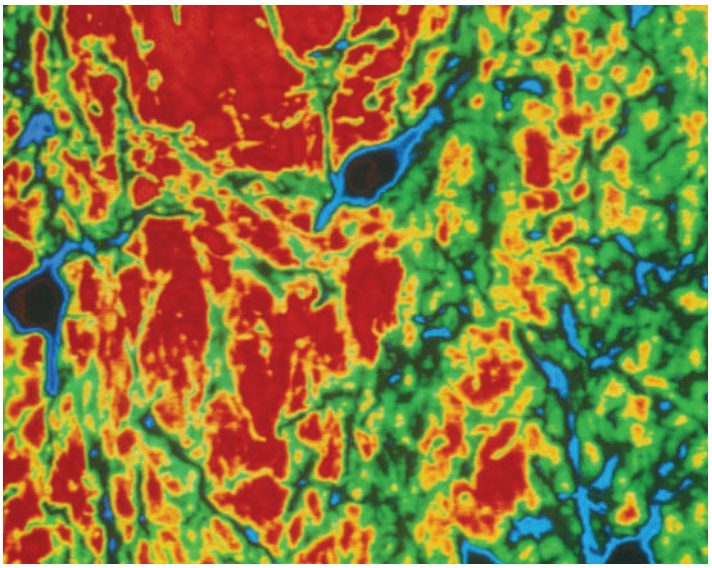
In this setting, Carlsson's ideas went down like a stone. The meeting unceremoniously rejected his interpretation of his data and, to his mortification, the single comment in the discussions praising his work was excluded from the symposium book. "The Ciba meeting might have been the opportunity to tell the world how things really were, but there was uniform hostility from the community," says Iversen.

Carlsson stuck to his guns, and data began to amass to a point that made denial impossible. Later in 1960, for example, the Austrian pharmacologist Oleh Hornykiewicz published studies on postmortem brains from patients with Parkinson's disease, showing the absence of dopamine in the basal ganglia[4]. And a few years later, Annica Dahlström and Kjell Fuxe at the Karolinska Institute in Stockholm, Sweden, showed that in the healthy brain neurons in this region contain high levels of dopamine[5].

In 1961, neurologist Walter Birkmayer, working together with Hornykiewicz, injected the first Parkinson's patients with L-dopa to dramatic effect, allowing previously immobile patients to move freely[6]. Carlsson recalls the penchant of his collaborator Tor Magnusson for testing drugs openly on himself — something that would be regarded with horror today. Expecting to see some kind of neurological effects, Magnusson hooked himself up to intravenous dopamine, but, says Carlsson, "all we saw was emesis!" Many other aspects of practising science were different then. Fax and e-mail did not exist, and all papers were read in the library rather than on a computer. "We certainly wore white lab coats and we addressed secretaries and assistants more formally as Miss," Carlsson says.

Over the next few years, clinicians learnt to start patients on low doses of L-dopa to avoid side effects such as nausea and vomiting. They also noticed other intrinsic imperfections of the drug. After a few years of therapy, its effects temporarily, and unpredictably, switch off in

Arvid Carlsson collects his Nobel prize in 2000, more than 40 years after showing dopamine to be a neurotransmitter.



Degeneration of the nerve cells that secrete dopamine (red) leads to Parkinson's disease.

some patients. In severe cases, patients can suddenly become frozen and be stuck, immobile, for minutes or hours.

Neurologists were also starting to learn that dopamine is involved in much more than the control of movement. They noticed that high doses of L-dopa cause some of the symptoms of psychoses such as schizophrenia, and that the antipsychotic drugs used to treat schizophrenia can cause the same movement problems that are indicative of dopamine deficiency in Parkinson's. This led to the idea that schizophrenia could be related to a disturbance of dopamine neurotransmission in areas of the brain other than those involved in movement.

The molecular multitasker

Carlsson reasoned that antipsychotic drugs could be blocking dopamine receptors, and that neurons might be spitting out more and more dopamine to try to compensate for the blockage. He was right. This type of feedback control of neuronal activity is now a well understood, and critical, phenomenon in neurotransmission. And this way of thinking won Carlsson many fervent admirers. He is "just about the most creative, intuitive scientist I've met", says Solomon Snyder of the Johns Hopkins University in Baltimore, Maryland, who discovered opiate receptors in the brain.

Another side effect of L-dopa treatment is that patients may develop an irrational tendency to gamble. It is now well known that the neural pathways controlling behaviours such as motivation and feelings of reward pivot around dopamine. These pathways drive pursuit of food and sex — and are hijacked by addictive drugs and addictive behaviours such as gambling. In the 1960s, Carlsson was among the first to spot that drugs of abuse work by boosting dopamine transmission in particular brain areas.

Too little dopamine in one area produces Parkinson's, too much dopamine in others can cause psychoses. Over recent decades the importance and complexity of the dopamine system have mushroomed in scientists' eyes. Dopamine acts on many types of receptor, at varying levels and in different brain areas, and in concert with other neurotransmitters.

Another of Carlsson's legacies has been the development of dopamine stabilizers. These are dopamine-like molecules that have been chemically modified so that they activate dopamine receptors to only a certain degree, effectively constraining the level of dopaminergic activation in the brain to within the healthy range. The theory is that a stabilizer could compensate for lack of dopamine in Parkinson's without causing overactivation; or block the overactivity in schizophrenia without too much depletion. Several companies have dopamine stablizers in development, and one, aripiprazole, has been approved by the US Food and Drug Administration for use as an antipsychotic for schizophrenia and bipolar disorder.

When Carlsson reluctantly retired aged 66 — then the law in Sweden — he kept his research going by forming a company called A. Carlsson Research where, among other things, he developed a dopamine stablizer. Then, in 2000, Carlsson's work on dopamine and its control of movement was recognized with a share of a Nobel prize. "I won it 40 years after my discovery," he jokes. "Einstein won his some 20 years after his discovery. So I guess my work was twice as complicated as Einstein's."

> **"When I was an undergraduate student we were taught that there was no chemical transmission in the brain."**
> — Leslie Iversen

Until this time, Carlsson and dopamine had both followed stellar trajectories. But after the Nobel prize, Carlsson's luck began to falter — and the uncompromising side of his nature, which had served him so well in his scientific career, failed to help. Just a few months before winning the Nobel, Carlsson was voted off his own company's board of directors. And in the next few years, plans for his namesake institute also went awry.

Difference of opinion

The Arvid Carlsson Institute was launched in November 2004 with SKr630 million (US$92 million) funding over five years — a tribute from Gothenburg University and the regional authorities to the city's only Nobel prizewinner. Its original mission was to promote health care and neuroscience research in the region, and Carlsson was named honorary chairman of its developmental council. But disagreements began almost immediately about how the money should be divided up. Carlsson wanted a significant proportion to support his field of neuropharmacology, but others argued it should go to neuronal stem-cell research. Scientists decline to discuss details of the arguments, but the hostilities became so bitter that the institute was dissolved in April 2006. Carlsson's daughter Maria, who is also a neuropharmacologist, receives a small proportion of the funds — too small in the opinion of her father, "given that the research was stated to be done to honour my contributions to science".

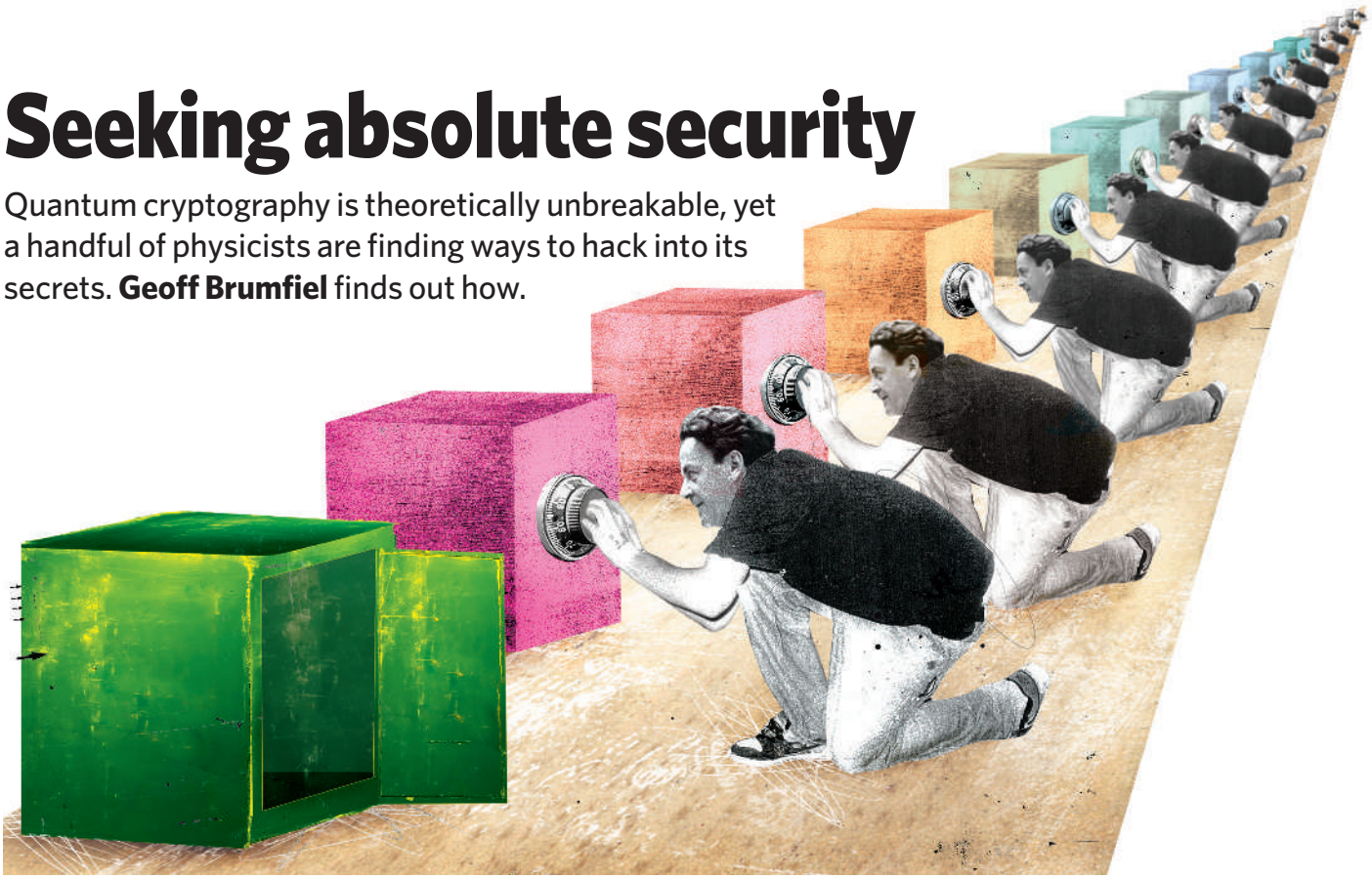
Carlsson's contributions to science continue. He is closely involved in his daughter's work and they sometimes publish together, although he wishes he could still work in the lab. He also jets around the world to meetings. Much in neuroscience has changed during Carlsson's time, but he believes at least one thing has remained constant. "When I started my career, the most important generator of science was the human mind," he says. "This has probably not changed much during the past half-century." ■

Alison Abbott is *Nature*'s senior European correspondent.

1. Sacks, O. *Awakenings* (E. P. Dutton, 1973).
2. Carlsson, A., Lindqvist, M. & Magnusson, T. *Nature* **180,** 1200 (1957).
3. Bertler, Å. & Rosengren, E. *Experientia* **15,** 10–12 (1959).
4. Ehringer, H. & Hornykiewicz, O. *Klin. Wochenschr.* **38,** 1236–1239 (1960).
5. Dahlström, A. & Fuxe, K. *Acta Physiol. Scand.* **62** (Suppl.), 232 (1964).
6. Birkmayer, W. & Hornykiewicz, O. *Wien. Klin. Wochenschr.* **73,** 787–788 (1961).

Seeking absolute security

Quantum cryptography is theoretically unbreakable, yet a handful of physicists are finding ways to hack into its secrets. **Geoff Brumfiel** finds out how.



D. ALLISON; L. DUNCAN/UNIV. ROCHESTER/AIP EMILIO SEGRE VISUAL ARCHIVES

On an otherwise quiet Saturday evening in 1946, Frederic de Hoffman briefly thought he had lost the secrets of the atomic bomb. De Hoffman was a physicist at Los Alamos National Laboratory in New Mexico. As part of his job, he kept the design for the weapon in nine filing cabinets in his office. When he came into work and opened one, he found an enigmatic note: "When all the combinations are the same, one is no harder to open than another — same guy."

De Hoffman thought that the 'same guy' was the person who had tried to break into the lab's secure facility earlier that summer, but, as it turned out, the thief was standing next to him. It was Richard Feynman (pictured above), a leading quantum theorist who had a reputation as an incorrigible prankster. He had broken into de Hoffman's filing cabinets earlier that day to grab a few documents he needed to write a report.

Security has come a long way since 1946, but some things never change. Quantum physicists are now learning how to crack what is arguably the most secure form of data transmission ever conceived: quantum cryptography. This encryption method is theoretically unbreakable, but nevertheless, groups are finding weaknesses that may require rethinking the design of commercial systems. The work, says Seth Lloyd, a physicist at the Massachusetts Institute of Technology (MIT) in Cambridge, is a cautionary tale. "Nothing is unassailable," he warns.

Quantum cryptography uses the fundamental laws of physics to encode information in the quantum states of particles, usually photons. Most existing systems use a protocol known as Bennett–Brassard 1984, or BB84, which generates a secure quantum 'key' that can be used to encode messages sent between parties. BB84 works like this: the sender transmits an encoded key by polarizing single photons along one of two axes — up and down or tilted at 45° — and sending them along a fibre-optic cable to the receiver (see diagram). The receiver then randomly chooses an up-and-down or tilted filter to read each photon. If the filter they choose is aligned with the sender's original polarization, the receiver will be able to read one bit of the sender's data, but if they choose the wrong alignment, then the photon, by virtue of quantum mechanics, will pass through the filter in a random orientation.

"Not enough attention has been paid to vulnerabilities."
— Daniel Gottesman

After the original key has been transmitted, the sender and receiver compare the filters they used for each photon. They throw away the random bits and keep the rest as part of a new, secure shared key, which is then used to encode a longer message sent over regular channels.

At first glance, the BB84 protocol may seem complicated and highly inefficient. But the method behind it means that it can't be intercepted without the sender and receiver finding out. Suppose an eavesdropper were to try to listen in on their conversation. As she reads each photon with her own two filters, she passes it along to the receiver, but not necessarily in the same orientation as it was originally sent. Therefore, when the sender and receiver compare their keys, they will find a lot more random bits created by the eavesdropper and they can immediately cut off their communication or try to send a fresh key through a different channel.

What sets the BB84 protocol apart from other forms of cryptography is that the code should be impossible to crack. Most of today's keys are encrypted with a mathematical technique that depends on large prime numbers. Security hinges on the idea that large numbers are hard to factor into primes, but there is no way to be sure of that assumption, says Daniel Gottesman, who studies quantum cryptography at the Perimeter Institute in Waterloo, Canada. "We don't think there's a way to do it on a classical computer in any reasonable amount of time, but there is no way to prove that," he says. By contrast, the security of BB84 and other quantum protocols hinges on the immutable laws of physics: "Given that quantum mechanics is correct, then we can mathematically prove that this idealized BB84 protocol is actually secure."

But if the idealized version of the BB84 protocol is secure, the real version can be anything but, according to Charles Bennett, a computer scientist at IBM Research in Yorktown Heights, New York. Bennett is one of the 'B's in the BB84 name, and he and other researchers built the first demonstration unit in 1989. In that very first quantum-cryptographic system, Bennett recalls, the polarization of the photon was switched by use of

a high-voltage power supply. "The power supply hummed differently depending on whether or not the voltage was being applied," Bennett says. "If you listened, you could hear it."

No escape from reality

Nobody was planning on sending state secrets over the experimental set-up in Bennett's office. But while showing that quantum cryptography could work, he and his collaborators inadvertently demonstrated something else: idealizations are often far from reality. "It's hard to ensure that any box that you build is entirely secure," he says.

Such real-world security is the key to moving quantum cryptography from the lab to the commercial sector, and it has been a slow process. BB84 protocols require sending single photons, but early technology often sent more than one photon at a time, raising the possibility that an eavesdropper could read one without disturbing the others. Single-photon systems became commercially available a few years ago, but they remain modest in their capabilities. Error rates can be high, data speeds slow, and they can only be transmitted as far as a single photon can travel along a commercial fibre-optic line.

Meanwhile, researchers are stepping up their attacks on quantum cryptography. The most scientifically sophisticated strike was conducted earlier this year by MIT physicists led by Jeffery Shapiro and Franco Wong[1]. The team stole information from a passing photon by entangling its polarization with its own momentum. This quantum-mechanical entanglement allowed the team to read about 40% of the key while leaving the polarization relatively untouched. But Shapiro and Wong both admit that an eavesdropper would be defeated just by increasing the length of the key. Commercial systems already use long keys to deflect such attacks.

Other vulnerabilities could be even more dangerous because they have been overlooked by theorists. For instance, theoretical physicists assume that the sender and receiver will have absolute control over their equipment. But the real world is less precise, says Nicolas Gisin of the University of Geneva in Switzerland. Gisin and his colleagues have shown that an eavesdropper could learn a sender's polarizations by shining a light down the fibre and into the sender's set-up[2]. Because the cryptographic protocol assumes that light will only come from the sender, it doesn't take into account such dirty tricks.

And still other attacks can take advantage of simple flaws in individual components. Earlier this year, Hoi-Kwong Lo and his colleagues at the University of Toronto in Canada showed that they could steal a commercial system's quantum secrets by exploiting a small defect in the receiver's photodetectors. The protocol under attack was different from BB84 in that it required two photons. The system switched on the two highly sensitive detectors only when it was expecting photons from the sender to avoid false alarms. But the detectors switched on at slightly different times. By delaying photons so that they arrived just as a detector was turning on or off, Lo showed that an eavesdropper could modify the measurement, which blinds the receiver to the eavesdropper's presence.

"It's hard to ensure that any box that you build is entirely secure."
— Charles Bennett

Not everyone agrees on how serious the threats are to commercial systems. "We are quite confident that our system will remain impervious," says Robert Gelfond, chief executive of MagiQ, a quantum-cryptography company in Somerville, Massachusetts. Gelfond says MagiQ's government and military customers regularly try to breach their systems' security. "They want to see it and test for themselves," he says. So far, MagiQ has not had to modify any of its cryptographic technology.

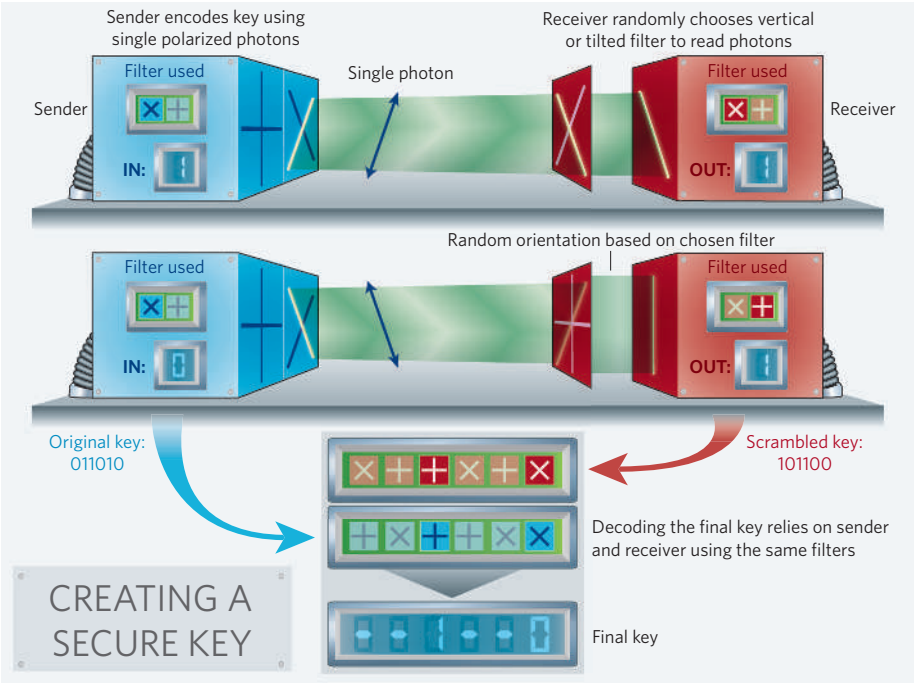
Weak spots

But others think that more needs to be done to ensure that the systems live up to their theoretical reputation. "There's been a lot of lip service," says Gottesman. "But I don't think enough attention has been paid to vulnerabilities." The researchers are more concerned with getting their set-up to work than they are with finding ways to cheat the system, he says.

That may be changing. Gisin thinks that the number of studies on potential attacks has risen over the past few years. "The entire field is getting more mature," he says. "Now is really the time to think about these things." As quantum cryptographic networks grow in size and complexity, they are also at risk from new kinds of attacks. Gisin would like to see the industry develop standards for detectors, transmission lines and other equipment that would help to close future gaps in security.

But there will always be a dirty trick to try, as the quantum-theorist-turned-safecracker Richard Feynman knew all too well. Feynman didn't rely on his theoretical brilliance to open the safes holding America's atomic secrets. He simply guessed the combination. He knew that his friend, a physicist, would undoubtedly choose a number he already had memorized, and the sly theorist got it on the second try: 27-18-28, the first six digits of the mathematical constant *e*. ■

Geoff Brumfiel is *Nature*'s physical sciences reporter in Washington DC.



1. Kim, T., Stork, I., Wong, F. N. C. & Shapiro, J. H. *Phys. Rev. A* **75,** 042327 (2007).
2. Gisin, N., Fasel, S., Kraus, B., Zbinden, H. & Ribordy, G. *Phys. Rev. A* **73,** 022320 (2006).
3. Zhao, Y., Fung, C.-H. F., Qi, B., Chen, C. & Loh, H.-K. Preprint at http://lanl.arxiv.org/abs/0704.3253 (2007).

Signing on

When you win a Nobel prize, you become much in demand. **Eric Sorensen** takes a look at how laureates decide which worthy causes to lend their name to.

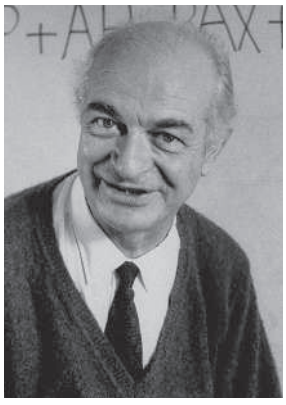
T. SPIEGEL/CORBIS

Half a century has passed since chemist Linus Pauling spearheaded one of the biggest petitions ever in science. More than 11,000 scientists, including 36 of Pauling's fellow Nobel laureates, signed on to call for the "ultimate effective abolition of nuclear weapons". The petition led to the first international attempt to control nuclear weapons — the Partial Test Ban Treaty. And on the same day in 1963 that the treaty went into effect, the Norwegian Nobel Committee announced that Pauling would receive the peace prize to go with his 1954 Nobel Prize in Chemistry.

Scientific petitions graced by laureates have become common tools of activism — clamouring to free the unjustly imprisoned and cure a myriad of perceived ills, from drug laws to inadequate research funding to nuclear proliferation. Having a Nobel laureate's name on a petition almost guarantees it extra attention: in a newspaper story's first paragraph, if not its headline.

The past year alone has seen laureates' signatures on petitions to make publicly funded academic research available for free on the Internet; decriminalize homosexuality in India; raise the US minimum wage; decry the Bush administration's alleged politicization of science; and restrict the US president's authority to order nuclear strikes against nations without nuclear weapons. And last week in Jordan, about 35 laureates gathered at the third Petra conference to discuss major world issues; it concluded with the launch of a US$10-million fund to bolster scientific projects in the Middle East.

As the high-powered scientific petition has grown, signature gathering has become its own industry. Leading the way is the watchdog group Union of Concerned Scientists in Cambridge, Massachusetts, whose 1992 World Scientists' Warning to Humanity on the environment was signed by about half of the living Nobel laureates in the sciences, for a total of roughly 1,700 researchers. Five years later, no fewer than 110 laureates signed the group's Call for Action on global warming.



Linus Pauling gained worldwide support for his petition to abolish nuclear weapons.

Politicians also routinely summon laureates — or at least their signatures — to their pet causes. During the 1996 presidential race, Bill Clinton had seven Nobel laureates backing his budget plan; his Republican rival Bob Dole had four. In 2004, George W. Bush's campaign mustered only six Nobel laureates to deride the tax plan of Democratic nominee John Kerry, which had the backing of 10 Nobel economists.

And that illustrates a fundamental problem with the Nobels: tease out the inner workings of matter, and you become a Nobel laureate; sign a petition, and you become a number. Roald Hoffmann of Cornell University in Ithaca, New York, won the 1982 Nobel Prize in Chemistry. He says that laureates become a sort of commodity from the moment he or she is asked if their name can be used. "It's a kind of detachment of the person from the subject," he says. "Do they really want to know what I think? Or do they just want my name?"

As the number of Nobel-signed petitions has risen, their value has decreased, says Peter Agre, who won the Nobel Prize in Chemistry in 2003 and is now vice chancellor for science and technology at Duke University School of Medicine, Durham, North Carolina. "The more you sign and respond to, the less valuable your service is," he says. Since winning the prize, Agre has signed petitions opposing the inclusion of intelligent design in science curricula and seeking leniency for an infectious-disease specialist charged with mishandling lethal biological agents. He also supported the candidacy of Kerry — along with 47 other laureates.

Great minds think alike

So how does a Nobelist, newly inundated with fame and requests, sort through the competing, well-meaning demands for his or her time and name? For Agre, it means looking at who else is already involved; if he sees other respected names on a petition, such as Harold Varmus from the University of California School of Medicine in San Francisco and winner of the 1989 prize for medicine, then he's in.

Nicholaas Bloembergen, emeritus of Harvard University in Cambridge, an honorary professor at the University of Arizona and winner of the 1981 prize for physics, says that he is asked to sign petitions half a dozen times a year. He signs about once a year, acting as a physicist on scientific issues such as federal funding for research and as a well-read citizen on social issues such as overpopulation. He was, for instance, one of 41 laureates signing a protest against the Iraq

D. CRAVENS/TIME LIFE PICTURES/GETTY IMAGES

war and one of 100 laureates to warn in 2001 that world security hangs on environmental and social reform.

Robert Solow, from the Massachusetts Institute of Technology in Cambridge and winner of the 1987 prize for economics, signs no more than half the times he is asked and he tries to stick to economics issues, recently advocating for a rise in the minimum wage.

Stairway to heaven

Solow says that: "The big difficulty is usually that you're asked to put your signature to some statement that someone else has written." If the statement is not in line with his thinking, he figures he has no business signing it; if he agrees with it in broad terms but not in specifics, he then asks if his disagreement with the author is minor enough. And he breaks little sweat over "general statements about peace and things like that. It's not my specialty but I read it over and figure when I get to the pearly gates, St Peter won't turn me away for favouring peace".

Aaron Ciechanover of the Ruth and Bruce Rappaport Faculty of Medicine in Haifa, Israel, and winner of the Nobel Prize in Chemistry in 2004, faces a lot of local demands for his attention. He is often asked and signs a few petitions — for instance, a petition calling on Israeli prime minister Ehud Olmert to open contacts with Syria and Hamas, or a call to the government in Sudan to stop the murder in Darfur. "I do not think that as a Nobel laureate my opinion is better or carries any extra weight than that of anybody else," he says. "Yet I am guided by my principles and conscience and am voicing my opinion on issues I think are important. At the end, it may add up."

Yet for all their celebrity, Nobelists seem to be decidedly weak instruments of social change. Leslie Gelb, Pulitzer prizewinning reporter for the *New York Times* and president emeritus of the Council on Foreign Relations, a non-partisan think tank based in New York City, says that he has seen petitions come and go over the past four decades. "I have not seen evidence that petitions change [the minds of] decision-makers," he says. Gelb has routinely asked people in power if they had seen petitions in full-page advertisements, and nine times out of ten, he was told they had missed it.

And perhaps that's not always a bad thing. "The big difference in life before and after you win a Nobel prize is there's nothing you can say that's so stupid that some magazine or newspaper won't print it," notes Solow. Others question the importance of some of these issues. "There is no petition so stupid that it cannot get at least a handful of signatures from Nobel laureates," economist George Stigler told a group of students at the University of Chicago in Illinois in 1970 — 12 years before he himself won the Nobel prize.

Even the best-intentioned petition can fail to measure up to its signatories' hopes. For instance, the 1992 'warning to humanity' encompassed a wide range of environmental issues, including ozone depletion, water pollution, declining fisheries, degraded soils, destroyed rainforests, species extinctions, overpopulation and poverty. Its release coincided with United Nations debate over actions outlined at the Earth Summit in Rio de Janeiro, Brazil, earlier that year. Yet few newspapers gave it more than a brief mention. "It is a very powerful and beautifully written document and it was just totally ignored," says Canadian biologist and broadcaster David Suzuki. "To me that is a stunning indictment of the kind of society we have that scientists are marginalized by the media," he says. Gelb, for his part, thinks that the laureates watered down their message by asking for too many things at once.

"You have to do what you can to get some attention."
— Dudley Herschbach

Still, laureates interviewed for this story would like to think that their support counts for something. The Nobel is a brand that many argue confers prestige and honour on petitions and their sponsors. "People like to put movie stars' and sports figures' names on petitions," says Philip Anderson of Princeton University in New Jersey, who won the 1977 Nobel Prize in Physics. "Is there any reason a sports figure would know more about famine or any other issue?"

J. EKSTROMER/AFP/GETTY IMAGES

Peter Agre's work on water channels won him the 2003 Nobel Prize in Chemistry.

To many, the consequences of remaining silent are too great to ignore. "The majority of Nobel prizewinners are willing to use the 'power of shame' to inform the public about developments that should not be accepted," says German physicist Klaus von Klitzing, director of the Max Planck Institute for Solid State Research in Stuttgart, and winner of the Nobel Prize in Physics in 1985.

History repeats

Dudley Herschbach of Harvard University and winner of the 1986 prize for chemistry, sees his activism as part of a long American tradition that stretches back to Benjamin Franklin, the eighteenth-century statesman and scientist. As an active laureate, Herschbach sits on the board of the Council for a Livable World in Washington DC. This political-action group was created by physicist Leó Szilárd, the person who first imagined a nuclear chain reaction and the leader of the Manhattan Project petition that failed to keep President Harry Truman from using the atomic bomb on Japanese civilians. Herschbach, for his part, worries about the 1,000 tonnes or so of weapons-grade enriched uranium that exists in the former Soviet Union, and the very real possibility that the uranium will fall into the hands of terrorists. "Things like that," Herschbach says, "you have to do what you can to get some attention."

But even he is well aware of a laureate's limits. Herschbach holds many of his field's highest honours, yet accepts that many people were more interested when he appeared on an episode of the television show 'The Simpsons'.

And perhaps it is just a symptom of democracy that a laureate may hold no more sway than any one else. "Each man counts for one," says economist James Buchanan, "and that is that." He should know. He won a Nobel prize. ■

M. DWYER/AP

Nobel prizewinners William Lipscomb (left), Robert Wilson (middle) and Dudley Herschbach make light of their status.

Eric Sorensen is a science writer in Seattle, Washington.

See Editorial, page 354.

Look at biological systems through an engineer's eyes

SIR — Your Connections series of Essays has taken some interesting looks at the interdisciplinary study of complex, dynamic systems (see www.nature.com/nature/focus/arts/essays/index.html). However, it has not featured a discussion of the physiological tradition of biological research, in which biological systems are analysed using reduced descriptions in much the same sense that an engineer uses a reduced description of an amplifier. An engineer is often not interested (to first order) in what is inside the box that produces gain, but studies the properties of the gain, its linearity, its frequency dependence and so on. A complete description of the structure of the amplifier is far less useful than a reduced description of its input–output relation, when the goal is to use the amplifier or connect it to other devices to make a system.

An engineer told that an unknown black box is an amplifier is rather like a biologist confronting an unknown biological system. Some structural knowledge is indispensable. Engineers would have a terrible time if they did not know which leads were power supplies, which inputs and which outputs. But the last thing an engineer would want to know is the complete circuit diagram, let alone the locations of all molecules or atoms in its resistors, capacitors and transistors. Successful investigation requires some (indispensable) knowledge of structure; but it requires many more measurements of inputs and outputs, under many conditions. Successful investigation also requires a good quantitative model of the system, called a device equation.

Physiologists have successfully analysed a large range of biological systems using this 'device-oriented' approach. For more than a century, medical students have used it to learn that the kidneys filter blood to make urine; the lungs transport oxygen from air to blood; muscles contract; sodium channels produce action potentials; and so on. Each device description in physiology — on each length scale from organ, to tissue, to cell, to organelle, to protein molecule — is associated with a device equation, just as a device description in engineering (for example, of a solenoid) is followed by an approximate device equation for its function, for example, its input–output relation.

No one knows which biological systems can be viewed productively as devices. No one knows how many of the unsolved complexities of biological research reflect problems of the reverse engineering of simple devices, and how many reflect the inherent complexity of biological systems. One can certainly imagine simple systems that are hard to investigate because of the paucity of experimental knowledge. Complex systems — for example, with many internal nonlinear connections like the integrated circuit modules of digital computers or, perhaps, the central nervous system — may not be easily analysed as devices, no matter how many experimental data are available. But it seems clear, at least to a physiologist, that productive research is catalysed by assuming that most biological systems are devices. Thinking today of your biological preparation as a device tells you what experiments to do tomorrow.

Asking the questions in this way leads to the design of useful experiments that may eventually lead to the device description or equation, if it exists. If no device description emerges after extensive investigation of a biological system, one can look for other, more subtle descriptions of nature's machines.

R. S. Eisenberg
Department of Molecular Biophysics and Physiology, Rush Medical Center, 1653 West Congress Parkway, Chicago, Illinois 60612, USA

Endowments are necessary for museums to thrive

SIR — You suggest that combining science and outreach might help museums succeed in achieving the funding they need for research, in your Editorial 'Museums need two cultures', News story 'Smithsonian looks beyond ousted boss' and News Feature 'Endangered collections' (*Nature* **446,** 583, 594 and 605–606; 2007). As the current and former heads of such institutions, we emphasize that this combination can succeed only if followed up with major fund-raising for endowments.

The slow growth in research funding is being countered with increases in the number of investigators. Moreover, cost recovery from grants never covers full institutional costs, and income from exhibitions barely covers the costs of the exhibitions themselves. Most museums, botanical gardens and related institutions are substantially under-endowed.

Recognizing this situation, the Visiting Committee for Smithsonian Science has called for a major campaign to build a new endowment of at least $1 billion. Without such a source of unencumbered funds, research and outreach in such institutions will starve. Institutions that have resisted this kind of fundraising, including the Smithsonian and the Academy of Natural Sciences, will continue to struggle financially.

Adequate endowment is particularly important for institutions with extensive and irreplaceable collections of biological specimens. It would be a tragedy if we did not capitalize on the knowledge we could gain by providing adequate funding for this research. Growing capital wealth in all parts of the world today means that there is a new opportunity to build the endowments for these unique institutions.

D. James Baker*, Jeremy A. Sabloff†, Peter A. Raven‡
*Visiting Committee for Smithsonian Science, 8031 Seminole Avenue, Philadelphia, Pennsylvania 19118-3915, USA
†University of Pennsylvania Museum of Archaeology and Anthropology, 3260 South Street, Philadelphia, Pennsylvania 19104-6324, USA
‡Missouri Botanical Garden, PO Box 299, St Louis, Missouri 63166-0299

Bright idea to improve prose but remain accurate

SIR — I agree with Cheryl Strauss ('Increasing prose quality by decreasing repetition' *Nature* **446,** 725; 2007) that it would be pleasing to find trimmer descriptions in the scientific literature. It is tempting to blame overuse of words such as 'increase' and 'decrease' on a linguistically repressive science culture.

Nevertheless, scientific communication aims to record information explicitly, leaving nothing to contextual interpretation, so that experiments may be repeated and verified. Unconventional description has the dual danger of being imprecise or too specific.

Take, for example, the suggestion of replacing 'increased' with 'brighter' or 'more intense' to qualify the word 'fluorescence'. All three expressions have explicit, exclusive definitions. 'Increased' fluorescence means that the number of photons emitted due to absorption of smaller-wavelength photons is larger than it was. If a sample is 'brighter', the luminous flux per unit area per unit solid angle has increased, scaled for human visual receptors. And if a sample is 'more intense', it might be fluorescing at the same rate into a smaller area. A scientist reading the article might make an incorrect assumption if the wrong qualifier is used. We must ensure that editing preserves precise scientific meanings.

I agree, though, that 'shorter mouse tails' is an improvement on 'mouse tails of decreased length' — no one is likely to assume they were shorter in time!

Brad Deutsch
Institute of Optics, Wilmot 121, University of Rochester, River Campus, Rochester, New York 14611, USA

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 500 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

COMMENTARY

P. GINTER/ESRF

Life scientists have in the past relied on physicists to build facilities such as the European Synchrotron Radiation Facility in Grenoble, France.

Laying solid foundations for Europe

European life-science infrastructure has been neglected for too long. The next generation of facilities needs better coordination and community support, argue **Iain W. Mattaj** and **Glauco P. Tocchini-Valentini**.

Biological research in Europe has benefited from many triumphs of 'small science'. Critical contributions made by individuals or small teams, such as Frederick Sanger's sequencing of insulin, or Erwin Neher and Bert Sakmann's studies of single ion channels, have advanced entire research fields. But biomedical sciences have now entered an era in which many projects require interdisciplinary approaches, extensive collaboration or even, as in the case of the Human Genome Project, industrial-scale equipment and staff. Many life scientists rely on access to complex equipment, specialized services or large collections of patient samples, mutant animals or biomolecular data. Marshalling support for these and other research infrastructures needs urgent action at various levels.

As a priority, life scientists need to organize themselves, and European member states must decide on their individual and collective strategies for selecting and funding infrastructure. These issues will be discussed at a meeting of European research ministers in Brussels, Belgium, this week and at the Fourth European Conference on Research Infrastructures in Hamburg, Germany, on 5 June (www.ecri2007.de). These meetings are part of an ongoing process that, over the next few years, could provide Europe with a long-term strategy on research infrastructure and initiate work on new projects.

By many measures, biology in Europe is healthy. To provide one illustration, the volume of data produced by life scientists has increased dramatically in the past 10–15 years, and will continue to grow. The biomolecular data resources of the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) are now consulted by more than 1 million users per year. Such rapid growth was not expected when the EBI was founded 14 years ago. Consequently, it is now generally accepted that advances in biomedical research need to be backed up with adequate and widely accessible infrastructure and facilities.

Yet support for European life-science infrastructures remains woefully limited. The EBI, for example, is funded at a much lower level and, because it depends on research grants for almost half its income, in a much less stable manner than its closest US counterpart (the National Center for Biotechnology Information). The lack of infrastructure support has negative consequences for Europe's capacity for innovation in the biosciences and for international competitiveness.

Why has life-science infrastructure not kept pace with research? In the past, the field has been content to piggy-back on the physics community, who built the synchrotron rings and neutron sources now used for structural biology. Alternatively, we have established infrastructures with inadequate and fragile funding mechanisms. In our view, it is time for European life scientists to ensure that their views reach and influence decision-makers. Without political support, new European-scale infrastructures will never become a reality, irrespective of how desirable or necessary they might be.

Gaps in the roadmap

Of course, different scientific fields have differing infrastructure requirements. Physics, energy and space research still need 'big machines' with large construction costs. The life or social sciences, in contrast, more often require archives or collections, or access to distributed specialist resources such as mouse-phenotyping facilities. For these infrastructures, annual running costs are substantial and can quickly exceed the construction budget. European and national funders are attuned to 'high initial, low running' infrastructure costs and although awareness is growing of alternative models, a change is needed in the funding mindset.

A recent example of oversight in European research funding is the European Research Council (ERC). The ERC is a new and important funding source because it recognizes the value of pan-European competition on the basis of scientific excellence. But funding of individual research projects, important as it is, is only one aspect of funding research. The ERC was never intended to fund research

infrastructure but, to be competitive, recipients of ERC funding will require access to state-of-the-art facilities and infrastructure. Research infrastructures are used by many different parts of the scientific community and, in our view, are a natural subject for European, rather than national, planning and execution.

European scientists do have an infrastructure roadmap constructed by the European Strategy Forum on Research Infrastructures (ESFRI). The ESFRI was set up in 2002 by European member states and the European Commission. Its mandate is to provide an ongoing forum for informal consultations on research infrastructure. The 2006 roadmap produced an excellent list of priority projects, but this is not a final list of projects that will be carried out and, crucially, it comes with no funding commitment.

The ESFRI roadmap consists of 35 infrastructures selected to satisfy the needs of broad sections of the scientific community. Each project is considered mature enough to receive support from European member states and the European Commission. Six of the 35 represent the biomedical sciences (see 'Six for Europe'), with the other 29 covering a broad range of research disciplines. It is heartening that the ESFRI discussions have, from an early stage, acknowledged the importance of biomedicine and concluded that European infrastructure in the life sciences is as important as that in the physical sciences.

Six for Europe

BBMRI: European Biobanking and Molecular Resources Infrastructure (*Austria)
EATRIS: European Advanced Translational Research Infrastructure (*Germany)
ECRIN: European Clinical Research Infrastructures Network (*France)
ELIXIR: European Life Sciences Infrastructure for Biological Information (*EMBL)
Infrafrontier: Functional Genomics in the Mouse as a Model of Human Disease (*Germany)
Instruct: Integrated Structural Biology Infrastructure (*UK)

*Organization or country taking the lead in proposals submitted to the Seventh Framework Programme.

EMMA

Infrafrontier will integrate existing mouse archives.

Discuss and lobby

Although the European Commission strongly endorsed the ESFRI report on its release in 2006, there is no commitment from the member states to support any specific infrastructure project. In the next Framework Programme (FP7), which runs from 2007 to 2013, the budget for research infrastructure is only 42% of that requested (and 3.4% of the total research budget). Because most of these funds are already earmarked to support existing programmes, little is available for the ESFRI roadmap. This is disappointing, but the FP7 is not the only means of funding the roadmap and there now needs to be serious discussion of the mechanisms by which any of the ESFRI projects might be realized.

"European infrastructure in the life sciences is as important as in the physical sciences."

It seems likely that every ESFRI project will be supported by a different combination of funders. There are FP7 funds available to support a planning phase for the 35 roadmap proposals, including the identification and recruitment of potential national funders, presumably from among the ESFRI member states. This means that every member state will need to initiate its own consultation process to decide on strategy and priorities. Life scientists must therefore take the initiative to lobby for those infrastructures that they need for their research. The decisions on the first projects are expected in the next year or two, and organizations that represent scientists, such as academies, research councils and scientific societies, should get involved at the national level, if they are not already.

European states receive 'structural funds' from the European Union to improve national infrastructure — from roads to universities — and according to Robert-Jan Smits, who leads the European Commission directorate responsible for research infrastructure, these funds can be used both to support the construction of new research infrastructures and for membership of existing European facilities. At this week's meeting of European research ministers it is hoped that the ministers can at least agree on how they will reach, and coordinate, future decisions on ESFRI projects.

What else can life scientists do to influence the process? The European Life Sciences Forum (ELSF, www.elsf.org) has made a start. Formed in 1999 and now supported by many European life-science organizations, ELSF encourages life scientists to participate in European science policy discussion and public debate. One of us (I.W.M.) has been president of ELSF since February. ELSF was instrumental in organizing grassroots support for the ERC, and last summer hosted a well-attended and lively debate on research infrastructures in Vienna, Austria. ELSF plans further public meetings and information gathering towards organizing specific infrastructure interest groups that will explore and pursue the needs of the life-science community.

No perfect model

Bringing multiple national systems together to reach decisions on the 35 projects on the ESFRI roadmap will be a massive task. All six life-science projects have submitted proposals to the FP7 (see 'Six for Europe'). They are supported by up to 50 partners including many funding organizations, although none has yet pledged any funding. Subject to evaluation, each project may receive €5 million (US$6.8 million) for a planning phase lasting several years.

Valuable models and sources of expert advice already exist at the European level, including the intergovernmental research organizations that currently build and operate research infrastructures (most of which are members of the EIROforum). Three of the seven largest intergovernmental organizations in Europe — EMBL, the European Synchrotron Radiation Facility and the Laue Langevin Institute — support life-science research. More recently, the European Mouse Mutant Archive (www.emmanet.org) has created a consortium of mouse facilities, which receives national funding but which coordinates the facilities' activities with the help of funding from the European Commission.

There is no perfect model for international research infrastructure, but intergovernmental organizations have historically been successful because they are legal entities with customized governance that allows independent operation but continuous external performance evaluation. However, such organizations take years to establish. Simultaneously setting up multiple entities with differing combinations of member states for various ESFRI projects seems to us impossibly difficult.

Achieving the next generation of European infrastructures will therefore require new legal and governance structures that can be more readily adapted to support successful ESFRI projects. In our opinion, these activities should eventually lead to the creation of an infrastructure counterpart of the ERC that will strategically evaluate and fund European infrastructure proposals. Until then, research infrastructure will continue to depend on individual member states for support despite the enormous difficulties in coordination that this entails. ■

Iain W. Mattaj is at EMBL, Meyerhofstr. 1, 69117 Heidelberg, Germany.
Glauco P. Tocchini-Valentini is at the National Research Council (CNR), Via E. Ramarini, 32, 00015 Monterotondo Scalo, Rome, Italy.

BOOKS & ARTS

Leading Los Alamos

To develop the atomic bomb, J. Robert Oppenheimer changed Los Alamos — and it changed him.

Oppenheimer: The Tragic Intellect
by Charles Thorpe
University of Chicago Press: 2007. 413 pp. $37.50

Catherine Westfall

Does the world really need yet another book about J. Robert Oppenheimer? The already high pile of Oppenheimer biographies has been elevated in the past three years by David Cassidy's book *J. Robert Oppenheimer* (Pi Press, 2004) and Priscilla McMillan's *The Ruin of J. Robert Oppenheimer* (Viking, 2005). There have also been collaborative studies by Abraham Pais and Robert P. Crease (*J. Robert Oppenheimer*; Oxford University Press, 2006) and by Kai Bird and Martin J. Sherwin (*American Prometheus*; Alfred A. Knopf, 2005). But amazingly, Charles Thorpe's *Oppenheimer* still manages to provide a fascinating new perspective.

Why have so many scholars tried to put together an account of Oppenheimer's life? Perhaps it is simply because the pieces are so intriguingly hard to mesh. In his younger years, Oppenheimer was a master of intellectual abstraction, an early expert in quantum mechanics who was also drawn to Sanskrit and communist politics. At Los Alamos he impressively managed the effort to build the first atomic bombs, making him a hero both inside and outside science. Although initially a strong advocate for using those weapons, after the Second World War he expressed qualms about developing the hydrogen bomb. He then precipitously lost power and respect, ensnared by McCarthy-era anti-communist politics and by his own testimony against friends at a highly publicized hearing that led to the revocation of his security clearance and the end of his government career. Oppenheimer continued to serve as director of Princeton's highly prestigious Institute for Advanced Study, a post he assumed in 1947. His speeches suggested that he felt guilt, but not regret, for ushering in the atomic age.

Like other biographers, Thorpe argues that Oppenheimer's contradictory behaviour arose from a poorly formed and therefore malleable self-identity. What's new here is a precise and compelling description of how Oppenheimer's Los Alamos persona was forged by wartime circumstances and the Los Alamos community. To succeed in its grim mission, Los Alamos needed a certain type of leader, and Oppenheimer nimbly fit himself to the role, becoming the intellectual, moral and social centre of gravity for the constellation of scientific and engineering problem-solving. Thorpe argues that just as Oppenheimer created Los Alamos, so Los Alamos created, or at least reconfigured, Oppenheimer.

AMERICAN INST. PHYSICS/EMILIO SEGRE VISUAL ARCHIVES/UNITED PRESS INTERNATIONAL

When worlds collide: J. Robert Oppenheimer had to work closely with the military at Los Alamos.

This approach might have resulted in sociological, postmodern sophistry. Instead, it helps to mesh apparent disconnections. For example, the congeniality that linked Oppenheimer and army general Leslie Groves, despite their divergent backgrounds and styles, now makes sense. At Los Alamos, Oppenheimer did his best to adapt academic tradition — with its leisurely pace and emphasis on continuously advancing knowledge for its own sake — to fit the job of wartime weapons-building, with its requirement to engineer rapidly using approximate knowledge. And because completing this military mission hinged on exploiting scientific expertise, Groves was willing to alter military tradition along quasi-academic lines to get that vital knowledge.

The book also shines new light on Oppenheimer's leadership. Thorpe is at his best when skilfully weaving quotations from the myriad of Los Alamos accounts and his own interviews, blending voices from oft-quoted scientists, seldom-included wives and largely forgotten military technicians. These accounts vividly describe how Oppenheimer acted as a mediator and buffer between the academic and military traditions, calmly soothing fears, easing moral concerns and lighting the way with his own keen intelligence. So, instead of complaining or fighting among themselves — and instead of second-guessing their mission after Germany surrendered — the diverse staff worked cooperatively under difficult conditions to solve hard technical problems on a tight schedule. The testimony itself strongly supports Thorpe's contention that the Los Alamos staff had a hand in shaping Oppenheimer's wartime persona. Oppenheimer, according to this testimony, was god-like, smarter and more noble than any human could be, a man much too good to be true. This larger-than-life persona was tailor-made for wider export, so it is not surprising that this version of Oppenheimer was embraced by the public in the heady days after the atomic bombs ended the Second World War.

By so clearly presenting the falseness of Oppenheimer's wartime persona, Thorpe

sets the stage for understanding why Oppenheimer later fell from grace: the gravity of postwar reality made the fall inevitable. However, Thorpe's analysis of the postwar years is much less impressive than his wartime study. The problem is that Oppenheimer was positioned to shape and be shaped by the compact, insular, war-focused Los Alamos, but the same was not true for postwar society. Thorpe tries to argue that Oppenheimer's experience in this period extends to all scientists — that work on the bomb joined science and the national security state together, leaving scientists compromised. Maybe they were compromised (as others have argued), but what happened to Oppenheimer cannot be seen as typical; he was too eccentric and his experience was unique. The Los Alamos portrait is apt, in fact, because it shows the precise relationship between a quirky leader and an odd community under unusual circumstances.

Understanding the evolving and complex relationship between scientists and the national security state requires a much wider focus than Oppenheimer's life. Indeed, understanding Oppenheimer's life in this postwar period requires a wider focus than McCarthy-era politics. Surely he was strongly influenced by his personal life, a subject Thorpe glosses over. Here, Thorpe lags behind the competition. The books by McMillan and by Pais and Crease provide a superior explanation of the security hearings, and those by Cassidy and by Bird and Sherwin provide a more comprehensive account of the entirety of Oppenheimer's life. Nonetheless, Thorpe's book provides the best perspective yet for understanding Oppenheimer's Los Alamos years, which were critical, after all, not only to his life but, for better or worse, the history of mankind. ■

Catherine Westfall is laboratory historian, Argonne National Laboratory, Argonne, Illinois 60439, and a visiting associate professor at the Lyman Briggs School of Science, Michigan State University, East Lansing, Michigan 48824, USA.

Safety first

Intervention: Confronting the Real Risks of Genetic Engineering and Life on a Biotech Planet
by Denise Caruso
Hybrid Vigor Institute: 2006. 272 pp. $17.95

Allison Snow

In *Intervention*, Denise Caruso challenges scientists to do a better job of evaluating the safety of genetically modified organisms (GMOs) and communicating unbiased findings to the public. Caruso, who founded the non-profit Hybrid Vigor Institute, examines with a healthy dose of scepticism the recent history of the regulatory policies affecting biotechnology in the United States. How, for example, can the Department of Agriculture simultaneously promote biotech research and agribusiness while also protecting the public and the environment from possible harm? In a broader context, how can the science of genetic engineering move forward and benefit society with sufficient oversight to prevent disasters? Caruso's answer is that we need to develop more transparent and democratic methods for incorporating scientific evidence in formal risk analysis and public policy.

One of the major strengths of the book is its accessibility to a general audience. Caruso, a former journalist, describes dry topics such as RNA interference and the US Coordinated Framework for Regulation of Biotechnology in terms that entertain the reader with wry humour and an appreciation for the absurd. In her view, molecular biology has "the whiff of the Holy Grail", and if you question the experts who promote GMOs, "you'll generally get a scorching look of suspicion". Sadly, many of the experts and industry representatives whom she targets are unlikely to read the book, although they should. I disagree with many of Caruso's conclusions, but I appreciate her thesis that the immense power of molecular biologists to redesign living organisms requires more scrutiny with each passing year. The release of certain transgenic crops, trees, fish, insects, viruses and bacteria into the environment could do much more harm than good. Malicious uses, such as designing transgenic organisms for bioterrorism, provide a worst-case scenario.

ALAMY

Warning sign: the number of genetically modified organisms released into the environment could increase rapidly.

Despite her frustrations with scientific experts, Caruso is respectful of reports from the US National Academy of Sciences that deal with GMOs and risk assessment, especially the National Research Council's *Understanding Risk*. She notes that these reports offer constructive recommendations that have yet to be implemented, either in the United States or elsewhere. At the same time, she boldly challenges a fundamental tenet of the reports and all US regulatory policy, namely the notion that risk assessment should focus on the actual products or traits of GMOs case by case, rather than the engineering process used to obtain them. Mainstream scientists and regulatory agencies typically assume that the use of recombinant DNA is irrelevant to risk assessment because genetically modified products are carefully examined for unintended effects before deregulation. In other words, genetically modified products such as insect-resistant maize are "generally regarded as safe" unless proven otherwise. Likewise, the US Food and Drug Administration accepts the idea that expert opinion and a battery of lab tests are sufficient to prove that genetically modified food is "substantially equivalent" to its non-transgenic counterparts (which can also have genes that cause unwanted health effects).

Caruso develops a series of worst-case scenarios, some of which are rather far-fetched, to illustrate why the doctrine of "product, not process" may be wrong. She contends that the overconfident zeal of molecular biologists and strong economic pressures to rush genetically modified products into global markets have squelched legitimate scientific enquiry into the possible risks, including dangerous outcomes that could be inherent to any GMO. What if gene splicing causes novel interactions between native and introduced DNA in a given transgenic crop, resulting in subtle yet harmful effects on human health? Have government agencies and the biotech industry fully examined this possibility? No, she asserts, because "our appointed arbiters of risk" are not willing to discuss the limitations of their knowledge. Moreover, she makes a convincing argument for why it is exceedingly difficult to predict the long-term and large-scale effects on human health and the environment of intentionally produced genetically modified traits. Recognizing that all new technologies bring a mixture of risks and benefits, she then discusses the advantages of allowing ethicists, social

scientists, environmental scientists and others to participate in discussions about risk analysis and public policy. Her point that risk assessments involve value judgements beyond the realm of pure science is well taken.

Countering the industry's spin on the benefits of biotech in both developed and developing countries, Caruso focuses on the dark side of genetically modified crops. Her book echoes many of the themes from Deborah Koons Garcia's *The Future of Food*, a documentary film that attacks all unsustainable and chemically intensive modern agriculture (www.thefutureoffood.com). Unfortunately, Caruso's reliance on websites, unofficial reports and news media for citations means that many of her findings should be checked for accuracy and context. For example, she describes reports that the cultivation of genetically modified crops has already harmed soil organisms, created superweeds, contributed to severe economic hardships, and made people and livestock sick by increasing their exposure to herbicides. Similar criticisms could be made of some non-transgenic crops. Indeed, certain transgenic crops offer greater health benefits than their conventionally produced counterparts. Regarding Terminator technology for producing non-viable seeds, she states that "critics fear that these plants would irreversibly spread their sterility to non-transgenic crops and across species to other plants by contamination". However, these yet-to-be-released crops would not bear viable offspring and so could not spread their genes through reproduction.

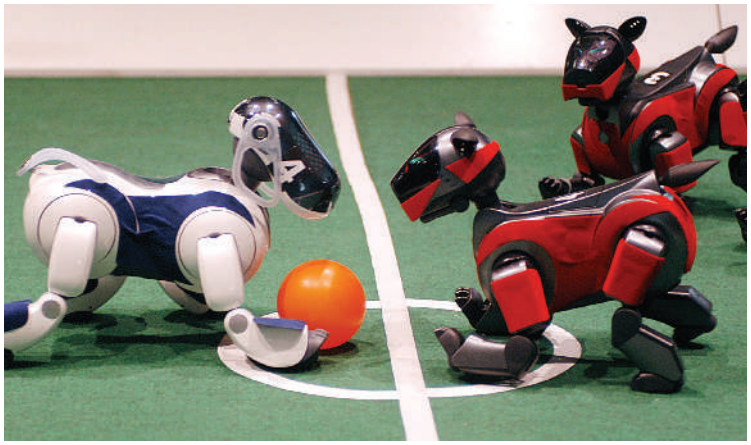
Caruso's fears that transgenes could spread willy-nilly to the genomes of unrelated plants and animals — and even to humans — are overly paranoid because distantly related multicellular organisms are not capable of interbreeding. In a flight of hyperbole, she states: "Billions of transgenics have already been released into the marketplace and thus into our food, our water and the air that we breathe, breeding and exchanging their genetic material with each other and with us." But perhaps today's hyperbole could be a prelude to the future, if GMOs are released indiscriminately around the world.

Reading *Intervention* made me more aware of the value of confronting uncertainty in the complicated process of assessing risks and benefits. This ambitious and engaging book does a good job of defending the layperson's frustrations and concerns about genetically modified organisms. ■

Allison Snow is in the Department of Evolution, Ecology and Organismal Biology, Ohio State University, Columbus, Ohio 43210, USA.

MORE ON GENETIC ENGINEERING

GM Crops: The Impact and the Potential/ Seeds for the Future: The Impact of Genetically Modified Crops on the Environment
by Jennifer A. Thomson
CSIRO Publishing/University of Cornell Press



MESSE BREMEN

Playing ball: robots must collaborate if they are to succeed at team sports.

Intelligence in a changing world

How the Body Shapes the Way We Think: A New View of Intelligence
by Rolf Pfeifer & Josh Bongard
Bradford Books: 2006. 394 pp.
$39.95, £25.95

Hiroaki Kitano

The study of intelligence, once dominated by biologists, has for decades been a focus for computer scientists. The question of whether a machine can be intelligent is as old as computers themselves, but was thrust into the spotlight ten years ago when IBM computer Deep Blue beat world champion Garry Kasparov at chess. Whether it really showed intelligence as we know it is still a matter for debate, but researchers have identified that 'intelligence' in the context of chess depends on having a huge database, computing power to search for moves and the ability to learn from past games to obtain a 'goodness score' for each possible move.

However, there are clear differences between the way humans and computers play chess: a chess computer, unlike a human, does not have a body to enable it to interact with its environment, for example. This distinction differentiates two views on intelligence. One view is that intelligence is independent of the body and is unaffected by its existence, shape and function. The other view is that intelligence is contained within a physical body and that the body shapes the mind, an idea often referred to as physical embodiment or the presence of a behaviour-based agent. There is increasing recognition in the artificial-intelligence and robotics communities that the nature of the body significantly affects the mind, although it does not totally control it.

How The Body Shapes The Way We Think by Rolf Pfeifer and Josh Bongard provides an excellent perspective on how artificial-intelligence and robotics researchers have been tackling this issue. It is full of examples and thought-provoking discussions so that readers can easily follow some of the central debates on intelligence developed over decades. It also presents a chronological development of the field where appropriate.

The major focus of this book is to discover the design principle of an intelligent agent that has a physical and mobile body, has a high degree of autonomy, interacts with its environment and exhibits a broader range of behaviours than those single-task chess computers. It is not a book about how the body of an existing life form shapes its own mind, so there are only limited references to biology and neuroscience. Nevertheless, there are several parallels between artificial systems and biological systems. In one of the design principles, the authors point out the importance of redundancy, which also applies to biological systems. Some of these commonalities between artificial and biological systems can be seen as system-level principles that seem fundamental to a system's ability to exhibit intelligence, at least to an observer's eye.

One salient difference between the intelligent agents discussed in this book and traditional artificial-intelligence systems, as represented by chess computers, is the contextual thickness of system behaviours. Many of the robotics systems discussed in the book can cope, at least to some extent, with changes in the expected environment, tasks and other assumed conditions, whereas chess computers and other traditional artificial-intelligence systems are usually extremely fragile when faced with even a small change in such conditions. Behaviour-based robots should be able to perform almost flawlessly if the size of road or unevenness of terrain deviates from the initial assumption. However, the results will be catastrophic if a chess computer is given a chess board with nine rows and columns, rather than eight, as they are tuned specifically for the existing rules of chess. Imagine a thought experiment on a chess game between a behaviour-based system and an existing chess computer. The chess computer would be unbeatable with the defined rules, but if the rules were modified the behaviour-based system may do better.

The authors discuss learning, development

and evolution, an approach inspired by, but not identical to, the biological one. These are methods designed to cope with complex, open and dynamic environments in which predefined rules cannot properly define behaviours because of the overwhelming complexity and difficulty in describing the situation. So artificial agents have to evolve, undergo development and learn how to behave. The authors also discuss the social aspect of multiple agents, or collective intelligence. How do agents with different expertise collaboratively accomplish complex tasks of the sort seen in real society or in sports such as football? These features are considered to be critically important for a multi-agent robotics team to play football in the RoboCup (www.robocup.org). The book beautifully illustrates the development of ideas, why we need these ideas, and what the issues are.

The book focuses on artificial agents, but with a lot of inspiration from nature. This reflects the synthetic approach to understanding: by building intelligent systems we can reach a deeper understanding of intelligence in general. Biologists often criticize such approaches as not being faithful to the biological processes that inspired them. However, researchers in artificial intelligence and robotics are, for example, trying to find out how to design an aircraft inspired by bird flight, instead of replicating a bird itself. It was a huge step in engineering when bird flight was decomposed into thrust and lift, and re-implemented by fixed wings and engines to create modern aircraft. Memory, computing and learning have been similar elements of intelligence in chess machines. The grand question behind this book is the search for such essential ingredients for intelligence in an open, dynamic environment. ■

Hiroaki Kitano is director of Sony Computer Science Laboratories, 3-14-13 Higashi-Gotanda, Shinagwa, Tokyo 141-0022, Japan.

SCIENCE IN CULTURE

Surrealism bites back

Sink your teeth into Jean Painlevé's nature films at an exhibition in London.

Martin Kemp

The success of nature documentaries on television rests on their ability to show us the wonders of the natural world. Intimate scenes such as a shimmying sea urchin embedding itself in sand no longer surprise, despite the unfamiliarity of the subject, thanks to the skill of the film-makers.

Rather more surprising is the fact that the first sustained use of film technologies to create moving images of nature's secrets occurred not within the world of professional science, but in the artistic context of surrealism.

The surrealist movement, led by Salvador Dalí, Max Ernst, André Breton and André Masson, regularly exploited biomorphic shapes in strange, dream-like contexts. The organic realm came to be populated by the unfamiliar, the uncanny and the downright sinister.

The development of film techniques resulted in a kind of 'natural surrealism'. The pioneer was Jean Painlevé, who in collaboration with Geneviève Hamon carved out a long career from 1925 to 1975 as the master of the genre. He used it to serve both science and art cinema. Painlevé is now little known, but makes a welcome appearance in the exhibition 'Surreal Things', which can be seen at the Victoria and Albert Museum in London until 22 July.

Painlevé, whose mathematician father twice became prime minister of France, was trained as biologist. He then embarked on a professional career as an actor and became involved in the anarchy of Parisian avant-garde art. His films, all short documentaries, range from soberly descriptive (although always with a strange lyricism) to overtly whimsical and ghoulish. He worked with leading experimental composers to choreograph his work — and used the music of Louis Armstrong and Duke Ellington in his soundtracks.

LES DOCUMENTS CINEMATOGRAPHIQUES

Some of Painlevé's films, such as *Sea Urchins* from 1929, rely on descriptive commentaries and direct filming. The drama of these films comes from the camera work, printing and cutting. The beauty of the living and moving sea urchin is seductively captured, although the vivisection of a specimen to show its hugely efficient digestive tract induces a shiver of horror. As Painlevé claimed: "Scientific film requires study and instruction; it is not only a tool, but a grammar and an art."

Other films are self-consciously melodramatic. *The Vampire*, which is just 9 minutes long, opens with some close-up shots of voracious animals in action, including caterpillars that munch insatiably and blood-sucking leeches. It then stars a vampire bat feasting on the blood of a compliant guinea pig (see picture).

The bat, accompanied by Ellington's jaunty *Echoes of the Jungle*, unconsciously overacts in a way that would suit the most lurid of horror films — a parallel drawn in Painlevé's film by some short clips from F. W. Murnau's silent film *Nosferatu* (1922). Murnau's dark masterpiece of expressionist cinema, based on Bram Stoker's *Dracula*, had precociously used microphotography to portray a 'vampire polyp'.

Composed during the Second World War, *The Vampire* served as an allegory of Nazism. The implication is that Hitler's fascism infects the mind just as rabies from the blood-sucking bat infects the bodies of its victims.

Nature 'red in tooth and claw' has traditionally provided a source of drama for artists. The nature revealed by film in general — and by the underwater camera, time-lapse and micro-photography in particular — provided the alert surrealists with a disturbing vocabulary of form and motion that was well matched to their purposes.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK. His latest book, *Seen | Unseen*, is published by Oxford University Press.

OCEANOGRAPHY

Power of pull

As the complex interplay of forces in the ocean responds to climate change, the dynamics of global ocean circulation are shifting.

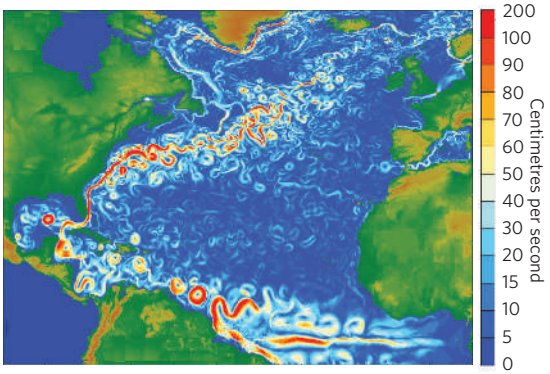
Martin Visbeck

Global climate change is edging the oceans towards a daunting future of sea-level rise, an increase in dissolved carbon dioxide and in acidity, reduced subsurface oxygen, and the possible loss of marine biodiversity and ecosystem functions. In all of this, large-scale ocean circulation plays a central role. Aside from affecting and to some extent being affected by these changes, the oceans are pivotal in global climate regulation. Any big shifts in this complex system present a considerable challenge to science and society.

In large-scale ocean circulation, often compared to a global conveyor belt, warm surface waters flow northwards from the equatorial Atlantic, giving up heat to the atmosphere. The cold, dense water sinks in the Greenland and Labrador Seas and returns to the Southern, Indian and Pacific Oceans as deep currents. This global overturning circulation is key to climatic stability because it contributes to more than half of the global ocean heat transport and allows many gigatonnes of CO_2 to be stored in vast amounts of deep ocean water, out of contact with the atmosphere. Global-scale differences in water density — built at the sea surface through solar heating, clouds and rain, and eroded by internal mixing caused by winds and tides — drive the movement of water. But what determines the vigour of global circulation: the 'push' of dense waters to depth in the polar regions, or the 'pull' of mixing those waters upwards throughout the ocean basins?

At Kiel University, Germany, in the late 1980s, we were focused on push — the sinking of cold, salty, dense surface waters in the North Atlantic, known as deep-water formation. We felt we were making exciting progress in understanding what drives the ocean conveyer belt. What we didn't realize then was that pull — the slow upwelling that arises from ocean mixing — was a crucial piece of this giant marine puzzle.

Push plays an obvious part in deep ocean circulation: if a deep, dense water reservoir grows in volume, its spread into all deep ocean basins will accelerate. But how does the water get from the surface to the depths? We began to see intriguing new direct measurements of vertical flows, capable of mixing fluid 'parcels' down to depths of 1 to 2 kilometres within a few hours, along with the first high-resolution computer models (basically upside-down cloud convection models) that could simulate this process in vivid detail. We learned that this deep convective mixing occurs sporadically, when high surface salinities come together with enough cold and windy winter days to make the surface water dense enough to let it plunge to the depths of the ocean. And it turned out that convection is also confined to a few key regions in the centre of the Greenland and Labrador Seas in late winter.



C. BÖNING & A. BIASTOCH (IFM-GEOMAR)

Simulation of near-surface ocean current speeds.

Push is also a key player in carbon sequestration. When surface and subsurface layers in those seas become cold and dense enough to cause intense vertical mixing, this water, replenished from the atmosphere with high levels of oxygen, CO_2 and other substances, is transferred to the depths. From there, a complex set of deep currents, most of which hug the western sides of the ocean basins, spreads the newly 'ventilated' water around the world.

Where does that leave pull? It took until 1994 for me to see its true importance. Confronted with simulations of one of the first global ocean circulation models, devised by Robbie Toggweiler at the Geophysical Fluid Dynamics Laboratory in Princeton, New Jersey, I was shocked. The model showed that artificially increasing wind speeds over the Southern Ocean enhanced deep upwelling, and the North Atlantic responded with an increase in overturning circulation.

Then I remembered Johan Sandström. In 1908, the Swedish oceanographer elegantly demonstrated how heating at depth is vital for maintaining a global-scale circulation in tank experiments. (After all, boilers are generally installed in basements so that the hot, less dense water flows upwards into a house's radiators.) To keep a density-driven ocean circulation going, you need vertical mixing of warmer waters into the ocean depths, which makes deeper water less dense, prompting upwelling — that is, pull.

Much remains to be discovered about ocean mixing. We do know it is largely driven by the wind and tide. Wind, for instance, creates surface waves that stir the upper layers of the ocean directly, and also induces surface friction that generates large-scale, deep-reaching currents. In their turn, these currents can create turbulent mesoscale eddies — ring-like flows with a radius of 10 to 100 kilometres. When wind- or tide-driven currents interact with undersea topography, steep subsurface internal waves are generated, some of which break just like waves on a beach to help mix density layers at depth. Tropical hurricanes and nekton — the huge schools of krill and small fish that create 'biological stirring' as they swim — are other possible sources of ocean mixing.

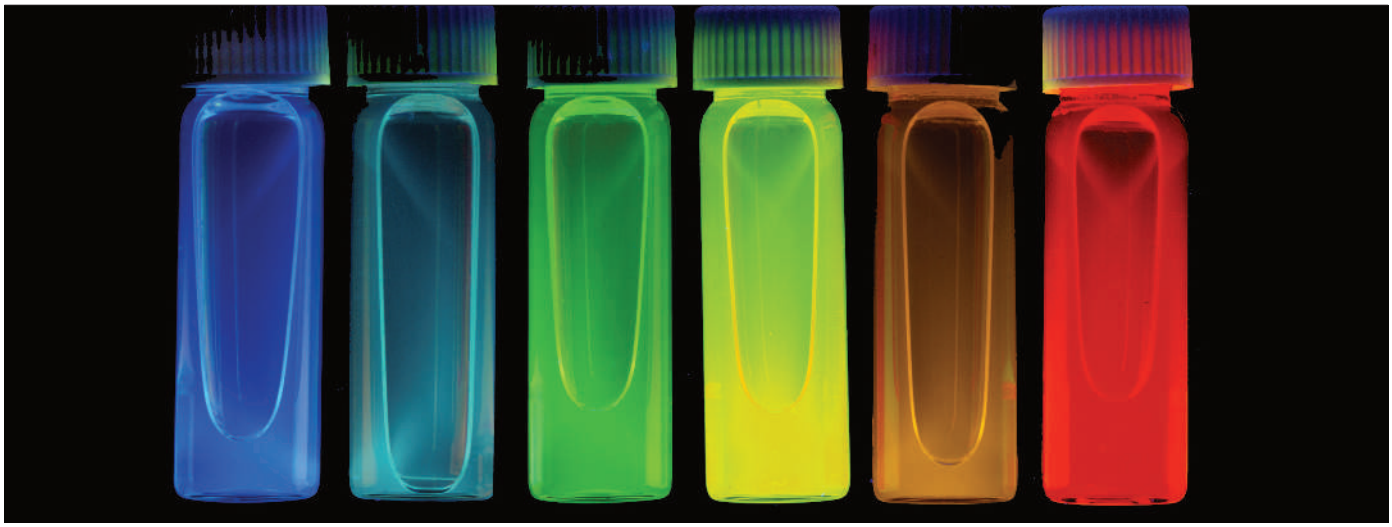
What does the future hold? By century's end, warming will cause more rainfall in polar regions, diluting the oceans along with meltwater from glaciers and the Greenland ice sheet. All this will lower ocean salinities and, together with warmer winter temperatures, reduce deep convection: the push will weaken. Current climate models predict that the Atlantic overturning will slow by 30%.

But over the longer term, shifts in pull will matter. Those same models suggest increased winds over the Southern Ocean and the possibility of stronger hurricanes. Both would boost ocean mixing and could, crucially, offset the trend towards slower circulation. More speculative is the role of any change in nutrient availability, which would affect the abundance of nekton and thus biological stirring. Global ocean observations, improved ocean models, smart brains and a multidisciplinary approach will all be needed to advance this frontier of science. ■

Martin Visbeck is at the Leibniz Institute of Marine Sciences at the University of Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany.

CONCEPTS

NEWS & VIEWS



F. FRANKEL, FROM *ON THE SURFACE OF THINGS* (HUP, OCT. 2007)

LASER TECHNOLOGY

Less excitement for more gain

Todd D. Krauss

In theory, semiconductor nanocrystals are highly suitable laser materials, not least because the colour of their light is tunable over a wide range. In practice, they are difficult — but not impossible — to deal with.

At the heart of any laser is a material that, when 'pumped' by an external energy source, amplifies a light beam. Colloidal semiconductor nanocrystals are particularly promising materials for a new and improved generation of lasers, but attempts over the past decade to construct efficient lasers out of them have proved frustrating: optical amplification was observed only under the most extreme and impractical of pumping conditions[1]. On page 441 of this issue, Klimov *et al.*[2] describe how they finely tuned the energy states of a nanocrystal to make pumping significantly easier. Their breakthrough could lower optical pumping thresholds by orders of magnitude, and so open the door for the practical use of colloidal nanocrystals in applications from telecommunications to medical diagnostics.

Colloidal semiconductor nanocrystals are crystalline particles typically between 2 and 10 nanometres in size[3]. When an electron in such a nanocrystal is promoted to an excited energy state, it leaves behind a locally positively charged region, a 'hole'. The electron and hole attract each other electrostatically, rather as the proton and electron in a hydrogen atom do, and form a particle called an exciton. Once created, this exciton in the nanocrystal decays into a photon extremely efficiently. Semiconductor nanocrystals are thus typically strong emitters of light.

In addition to their bright, robust fluorescence, as seen above, colloidal semiconductor nanocrystals have other unusual characteristics. Remarkably, the colour of the crystals' light emission can be tuned over a wide range simply by changing their size — an inherently quantum-mechanical effect[4]. They can also potentially emit in wavelength regions in the infrared, which current laser technology cannot easily reach[5]. Furthermore, an organic-molecule surface coating of the nanocrystals, such as trioctylphosphine oxide, can provide chemical reactivity and easy processability in solution, allowing them to be integrated without fuss into existing photonic-device configurations.

Despite all these advantages, the development of colloidal nanocrystal lasers has been severely hampered by the problem of achieving long-lived optical amplification. Optical amplification in a material occurs under a condition known as a 'population inversion', when — unusually — more members of a given population, say, of electrons in a semiconductor, exist in an excited state than in lower energy states. When a material that has a population inversion interacts with a photon of just the right colour, it will, in a process first described[6] by Albert Einstein, not absorb the photon, but instead will be 'stimulated' to emit two photons of identical colour to the incident photon.

For a successful laser, a photon in the initial laser beam must stimulate emission of two photons from the optical-gain medium before the population inversion is lost. For a typical nanocrystal, the lowest excited state can accommodate two excitons, so a population inversion, and successful lasing, require on average more than one exciton to be present per nanocrystal (Fig. 1a, overleaf). But because many excitons are confined to a minuscule volume in a nanocrystal, they interact strongly, causing one exciton to annihilate another in a process known as Auger recombination[7]. This process takes place incredibly fast, typically in much less than 100 picoseconds. Once it has occurred, any population inversion is ruined.

Achieving optical gain with nanocrystals therefore requires intense optical pumping with very short laser pulses, such that an avalanche of photons arrives at the nanocrystals before Auger processes have begun. Some limited progress has been made by replacing the nanocrystals with semiconductor nanorods[8], in which Auger lifetimes are slightly longer[9]. Nevertheless, for the past decade it has been widely assumed that lasers based on emission from excitons in colloidal nanoparticles would be made impractical by the excitons' short Auger lifetimes.

Klimov *et al.*[2] used cleverly engineered nanocrystals to split the exciton, funnelling the holes into a shell made of zinc selenide and

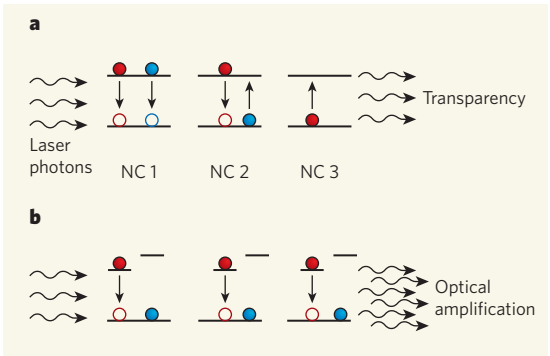


Figure 1 | Splitting the exciton. **a**, An ensemble of conventional semiconductor nanocrystals (NCs) is pumped to have, on average, one electron–hole pair ('exciton') (filled circles, electrons; empty circles, holes; downward arrows, stimulated emission; upward arrows, absorption). When illuminated by laser photons, double excitations (NC 1) contribute to gain, a single excitation (NC 2) means effective transparency (absorption and emission cancel), and lack of any excitation at all (NC 3) contributes to loss. Overall, this ensemble would exhibit optical transparency; because the doubly excited nanocrystals also decay very quickly into singly excited nanocrystals through the process of Auger recombination, much higher pumping rates are needed to obtain optical gain. **b**, In Klimov and colleagues' engineered core–shell nanocrystals[2], the energy required to create a second exciton (blue) is larger than that needed for the first. All singly excited NCs can contribute to optical amplification when lasing at the singly excited energy, thus significantly lowering the energy threshold for optical gain. In this case, the laser beam does not have enough energy to create a second exciton, so avoiding Auger processes and, potentially, allowing much slower pumping rates.

trapping the electrons in a core of cadmium sulphide. The relatively large physical separation of the electrons and holes means that a nanocrystal with one exciton has a different energy configuration from one with two excitons (Fig. 1b). In this case, a population inversion and optical amplification can occur for nanocrystals containing on average less than one exciton, thus completely avoiding the pitfalls of Auger recombination. Even more impressive is the fact that, under typical pumping conditions for optical gain, the lifetime of the excited state in these nanocrystals is almost 2 nanoseconds, 50 times longer than for typical colloidal nanocrystals[2]. The threshold for lasing depends inversely on optical-gain lifetime[10], and so it should now be possible to reduce this threshold by several orders of magnitude.

A hugely significant advance would be the integration of nanocrystal optical-gain media directly into a silicon optoelectronic device, such that the nanocrystals would be electrically (instead of optically) pumped. Such a scheme would find ubiquitous applications similar to the proliferation of electrically pumped semiconductor-diode laser devices, now found in everything from CD players to laser pointers to barcode scanners. The lifetime of a typical nanocrystal excited state is still far too short for these electrical pumping schemes: a typical semiconductor-diode laser has an excited-state lifetime of around ten nanoseconds[11]. But Klimov and colleagues' advances[2] bring this exciting possibility a lot closer. Stay tuned for further developments. ■

Todd D. Krauss is in the Department of Chemistry, University of Rochester, Box 270216, Rochester, New York 14627-0216, USA.
e-mail: krauss@chem.rochester.edu

1. Klimov, V. I. *et al. Science* **290,** 314–317 (2000).
2. Klimov, V. I. *et al. Nature* **447,** 441–446 (2007).
3. Alivisatos, A. P. *J. Phys. Chem.* **100,** 13226–13239 (1996).
4. Brus, L. E. *J. Chem. Phys.* **79,** 5566–5571 (1983).
5. Wise, F. W. *Acc. Chem. Res.* **33,** 773–780 (2000).
6. Einstein, A. *Phys. Z.* **18,** 121–128 (1917).
7. Klimov, V. I., Mikhailovsky, A. A., McBranch, D. W., Leatherdale, C. A. & Bawendi, M. G. *Science* **287,** 1011–1013 (2000).
8. Kazes, M., Lewis, D. Y., Evenstein, Y., Mokari, T. & Banin, U. *Adv. Mater.* **14,** 317–321 (2002).
9. Htoon, H., Hollingsworth, J. A., Dickerson, R. & Klimov, V. I. *Phys. Rev. Lett.* **91,** 227401 (2003).
10. Ebberly, J. H. & Milonni, P. W. *Lasers* (Wiley, New York, 1988).
11. Tang, C. L. & Olsson, N. A. *IEEE J. Quant. Electron.* **18,** 971–976 (1992).

CIRCADIAN RHYTHMS

Metabolic clockwork

Benedetto Grimaldi and Paolo Sassone-Corsi

The 'body clock' regulates the daily cycles of many physiological and metabolic processes, but just how is a mystery. New findings suggest that the cycling of energy metabolism is mediated by an activator of gene expression.

Intuitively, we all feel that the activities of our bodies follow cycles of repeated oscillations. Unmistakable examples are the sleep–wake cycle, the feeding rhythm and variations in body temperature and hormonal levels. Many of these cyclic oscillations are circadian (of around 24-hour periodicity), and are controlled by an interplay of numerous molecular factors. Such factors ensure the accuracy of the 'body clock', being organized in complex feedback loops that involve gene transcription and the events that follow it[1]. Liu and colleagues (page 477 of this issue)[2] provide a tantalizing interpretation of the molecular pathways implicated in the circadian control of energy metabolism, placing the transcriptional regulator PGC-1α in a strategic position.

The anatomical centre of the mammalian circadian clock lies within about 15,000 neurons in a region of the anterior hypothalamus in the brain called the suprachiasmatic nucleus. One unanswered question is how the suprachiasmatic nucleus directs the oscillatory nature of so many physiological and metabolic functions. The unexpected finding[3] that most peripheral tissues of fruitflies, zebrafish and mammals contain intrinsically independent pacemakers indicates the presence of a 'synchronization web' that coordinates timing in all the tissues. This concept, coupled with the striking notion that the transcription of at least 10% of all cellular genes oscillates in a circadian manner[4], underscores how profoundly the circadian transcriptional machinery influences a wide array of cellular functions.

PGC-1α is a transcriptional coactivator with an essential role in the maintenance of glucose, lipid and energy homeostasis[5]. It is highly responsive to a variety of environmental cues — from temperature to nutritional status and physical activity — and has been linked to several pathological conditions, including obesity, diabetes and neurodegeneration[5]. This pivotal contribution of PGC-1α to the coordinated regulation of metabolic pathways has acquired yet another dimension with Liu and colleagues' discovery[2] of its intimate links with the circadian clock.

Indeed, Liu *et al.* found that mice lacking the PGC-1α gene show abnormal circadian rhythms, as well as disruptions in the oscillatory pattern of the expression of 'clock' genes such as *Bmal1* and *Rev-erbα*. These PGC-1α-null mice also seem to have metabolic defects, such as aberrant body temperature and metabolic rate. The parallel with the outcome of targeted mutations of some circadian regulators in mice is striking: mutations of CLOCK and BMAL1 — two proteins that form a complex to stimulate the transcription of a variety of clock-controlled genes — also cause both an alteration in circadian rhythms and metabolic abnormalities in lipid and glucose homeostasis[6,7]. Thus, PGC-1α seems to occupy a privileged regulatory position, acting as an intimate link between metabolism and circadian control (Fig. 1).

What is the molecular mechanism through which PGC-1α influences circadian physiology? Liu and colleagues provide an interesting lead by showing that PGC-1α regulates the function of the ROR family of factors known as orphan nuclear receptors[2]. These factors oscillate in a circadian manner[8], and are involved in

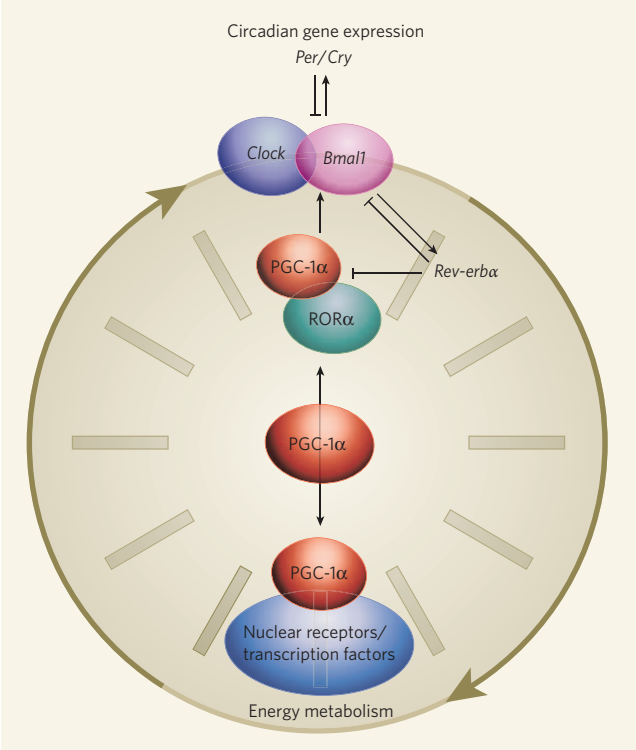


Figure 1 | An intimate link. Several genes are involved in regulating the circadian rhythm, including *Clock*, *Bmal1*, *Per*, *Cry* and *Rev-erbα*. The findings of Liu *et al.*[2] indicate that a coactivator of gene transcription, PGC-1α, has a central role in connecting the expression and function of these circadian-clock genes to the regulation of energy metabolism, which includes processes such as gluconeogenesis (the synthesis of glucose from non-sugar substrates), oxidative phosphorylation, oxidation of fatty acids and the biosynthesis of haem. This function of PGC-1α seems to be mediated by nuclear receptors, such as RORα, and other, unidentified, transcription factors.

regulating the clock machinery[9]. As RORα is required for the normal expression of the *Bmal1* gene — which counteracts the transcription-silencing activity of Rev-ERBα — and is essential for consolidating daily locomotor activity[9], it seems that PGC-1α might function directly at the core of the clock machinery. A couple of issues, however, deserve further exploration.

First, PGC-1α exerts its coactivator function by interacting with target genes, and recruiting to them various proteins that are involved in the remodelling of chromatin (complexes of DNA and histone proteins), especially histone acetyltransferase enzymes such as CBP, p300, P/CAF and SRC-1 (ref. 5). These enzymes act by adding an acetyl group to their substrate protein. As CLOCK has histone acetyltransferase activity[10], by virtue of which it can induce chromatin remodelling and establish a 'permissive state' for circadian activation of gene expression, it will be important to establish whether CLOCK and PGC-1α interact and whether CLOCK can acetylate PGC-1α.

Second, the clock proteins PER and CRY classically act as transcriptional repressors of the CLOCK–BMAL1 complex, thereby ensuring that the function of these activators of transcription is tightly regulated in a circadian manner. So it would be interesting to elucidate the specific function of PER and CRY in metabolic pathways, and any role they might have in the molecular control of PGC-1α activation.

Finally, the oscillatory expression of various clock-controlled genes is known to be under the control of cellular metabolic states. For example, changes in glucose levels can modulate and entrain circadian oscillations in cells grown in culture[11]. This observation is particularly interesting because the activation of transcription by CLOCK–BMAL1 can be affected by the intracellular levels of the NAD^+/NADH coenzyme, which is involved in glucose metabolism. Hence, it would be important to establish whether the contribution of PGC-1α to circadian regulation is dependent on NAD^+/NADH. This could be a physiologically relevant control switch, and might serve to differentiate the function of PGC-1α from that of other members of the PGC family.

In this context, it is intriguing to consider recent findings for PGC-1β, another member of the PGC family, but one with a different pattern of expression from that of PGC-1α. PGC-1β has now been implicated in circadian activity, as well as in the metabolism of the cell's energy generators, mitochondria, and in adaptive thermogenesis — the production of heat in response to changes in environmental temperature or diet[12]. Oxidative metabolism in mitochondria is already known to be tightly linked to PGC-1 activity[13]. Indeed, both PGC-1α and PGC-1β are known to have a role in the regulation of genes involved in the biosynthesis of haem (a non-protein component of certain proteins), as well as in ion transport, messenger RNA translation in mitochondria and cell respiration.

As abnormal PGC-1 activity will probably play an essential role in the pathogenesis of hyperglycaemia, insulin resistance and other disorders associated with impaired mitochondrial function and increased oxidative stress[5], it is tempting to speculate that some of these aberrant conditions could indeed result in, or depend on, altered function of the circadian clock. The possibility that these intriguing links could pave the way to novel therapies may then be quite realistic. ■

Benedetto Grimaldi and Paolo Sassone-Corsi are in the Department of Pharmacology, School of Medicine, University of California, Irvine, California 92697, USA.
e-mail: psc@uci.edu

1. Hastings, M. H., Reddy, A. B. & Maywood, E. S. *Nature Rev. Neurosci.* **4,** 649–661 (2003).
2. Liu, C., Li, S., Liu, T., Borjigin, J. & Lin, J. D. *Nature* **447,** 477–481 (2007).
3. Schibler, U. & Sassone-Corsi, P. *Cell* **111,** 919–922 (2002).
4. Panda, S. *et al. Cell* **109,** 307–320 (2002).
5. Lin, J., Handschin, C. & Spiegelman, B. M. *Cell Metab.* **1,** 361–370 (2005).
6. Turek, F. W. *et al. Science* **308,** 1043–1045 (2005).
7. Rudic, R. D. *et al. PLoS Biol.* **2,** e377 (2004).
8. Yang, X. *et al. Cell* **126,** 801–810 (2006).
9. Sato, T. K. *et al. Neuron* **43,** 527–537 (2004).
10. Doi, M., Hirayama, J. & Sassone-Corsi, P. *Cell* **125,** 497–508 (2006).
11. Hirota, T. *et al. J. Biol. Chem.* **277,** 44244–44251 (2002).
12. Sonoda, J. *et al. Proc. Natl Acad. Sci. USA* **104,** 5223–5228 (2007).
13. Lin, J. *et al. J. Biol. Chem.* **278,** 30843–30848 (2003).

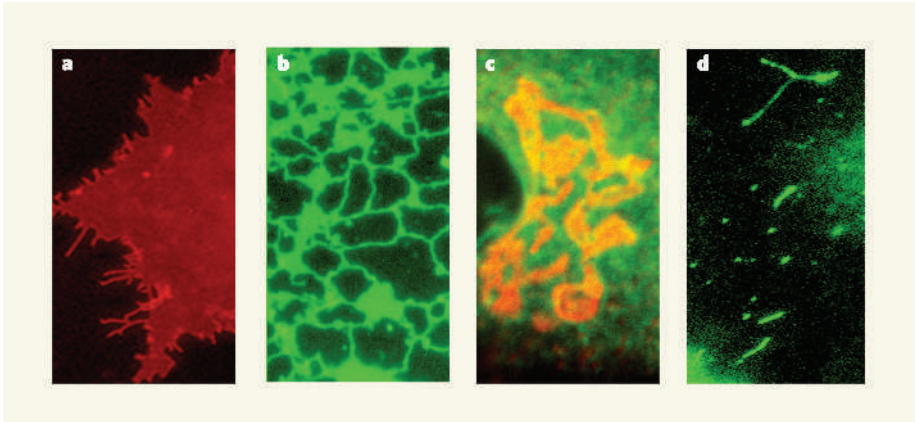
BIOPHYSICS

Bending over to attract

Michael M. Kozlov

What forces shape the membranes of the biological cell? A computer simulation indicates that it is the concerted effort of many proteins, mediated by the lipid bilayer that forms the membrane matrix.

Cell membranes are universal biological envelopes. The 'plasma membrane' forms the external boundary of a cell, separating its body — the cytoplasm — from the outside world. Other membranes wrap compartments within the cell, such as its nucleus and the 'organelles' responsible for essential cellular processes. Important among these are the endoplasmic reticulum and the Golgi complex, which mediate the synthesis of biological macromolecules. Membranes even form small carriers to transport substances within cells. These diverse physiological tasks are underpinned by one of the most striking and intriguing properties of cell membranes: their ability to adopt a great variety of shapes[1,2]. On page 461 of this issue, Reynwar *et al.* present[3] a state-of-the-art computer simulation of a model membrane, which they use to demonstrate its intrinsic ability to bend spontaneously into a biologically relevant shape.



K. HIRSCHBERG, TEL AVIV UNIV.

Figure 1 | Membrane twists and turns. The shapes of different biological membranes can be seen by labelling them with fluorescently tagged proteins. **a**, Plasma membrane (red). **b**, Endoplasmic reticulum (green). **c**, Golgi complex (red/yellow) and endoplasmic reticulum (green). **d**, Vesicular membrane carriers (green) for transport from the Golgi complex to the plasma membrane.

At one end of the scale of membrane shapes, plasma membranes adopt a configuration that is practically flat. Cylindrical membrane structures 30–100 nanometres in diameter characterize both the tubules of the endoplasmic reticulum and the membrane carriers operating between it, the Golgi complex and the plasma membrane. The membrane structures that mediate endocytosis (the uptake of material into the cell from outside) and exocytosis (the export of material from the cell), and those that contribute to intracellular trafficking, form spheres of 40–100 nm diameter. Finally, saddle-like shapes are found at membrane necks and junctions, as well as in fenestrations of the cisternae — the membrane disks that stack up to form the Golgi complex. All of these pure shapes coexist and intermix, and so form the beautiful, complex membrane architecture of the cell interior[4] (Fig. 1).

Living cells are essentially non-equilibrium systems in which all the components and structures, including the membranes, are in a state of constant flux. Membranes must therefore be able to undergo relentless shape transformations. These involve strong deformations of membrane surfaces and, in many cases, drastic rearrangements of membrane structures, leading to their fusion and fission.

Understanding how membrane shaping is regulated in time and space within the cell is a fundamental problem of contemporary cell biology. Biology itself specifies and characterizes the molecular players involved in membrane shaping. But both physics and mathematics have their part to play. Mathematics provides the tools to quantify the bent state of the membrane, through the idea of curvatures. And physics allows us to identify the factors and forces generating the membrane shapes and the transitions between them.

Cell membranes consist of lipids and proteins. Membrane curvature is therefore determined by the behaviour of one of these two classes of component, or by interplay between them. Lipids build up the membrane matrix, a bilayer film about 4 nm thick that, left to its own devices, tends to adopt an almost flat shape and resist strong bending. The large curvatures of intracellular compartments must therefore be generated by specialized proteins that are bound to the membrane surface or embedded in the hydrophobic core of the lipid bilayer.

Several classes of such proteins have been discovered, and the common structural features that allow them to generate and sense membrane curvature have been identified[1,5,6]. This is a necessary, but not sufficient, condition for understanding the shaping of cell membranes. A protein molecule can generate membrane curvature only locally, within a region comparable to its own size. But to form the intricate shapes found inside the cell, membranes must curve on a much larger scale.

This larger-scale curving requires the concerted effort of many proteins, which is possible only if the proteins come very close to each other and form domains on the membrane surface. Short-range binding forces are thought to stabilize these protein domains, but long-range attractive forces are needed to bring membrane proteins close enough for these binding interactions to come into play. The physics of this long-range attraction is a crucial open question.

One possibility is that the lipid bilayer matrix can itself mediate forces between the membrane proteins. This elegant idea was suggested more than a decade ago[7], and the theory behind it has since been extensively explored. There are two principal physical reasons for such forces. The first is straightforward: bilayer deformations produced by individual membrane proteins can overlap and influence each other, thus changing the elastic energy of the membrane and leading to an inter-protein force. The second reason is more subtle. Any membrane protein locally restrains thermal undulations of the lipid bilayer. Such undulations are favoured entropically, and so this straitjacket increases the overall free energy of the bilayer. Neighbouring proteins collaborate in restricting the membrane undulations and reduce the total energy costs, yielding an effective protein–protein interaction.

Most of the previous analyses showed that, in the biologically relevant cases, deformation-related forces are repulsive and should in fact prevent, rather than promote, the formation of protein domains. The restriction of undulation, by contrast, leads to an attractive force favouring the formation of protein domains. The outcome of the interplay between these attractive and repulsive forces has remained unclear. Both have the same dependence on the protein–protein distance, and their absolute values differ only by coefficients with similar values. But even when the attractive force is considered on its own, its predicted value seems too weak to drive domain formation.

The work of Reynwar *et al.*[3] could break this impasse and lead to a new understanding of membrane shaping by proteins. The authors' computer simulation modelled the behaviour of an ensemble of curved particles inserted into or adhered to a lipid-bilayer matrix. The simulations showed unambiguously that all types of membrane inclusion had a tendency to form domains: in other words, the net interaction between the proteins must be attractive. Moreover, the attraction between proteins with sufficient intrinsic curvature was strong enough to cause them to cluster. This implies that the interaction energy of the particles is considerably greater than the thermal energy, and so the attraction overcomes the entropically driven tendency of the particles to distribute evenly across the available area of the lipid bilayer. The membrane fragment covered by this cluster adopted the shape of a spherical bud, with a curvature similar to that found in membranes of real cells.

If the results of these simulations are universal, rather than limited by the specific model used, the attraction mediated by the membrane's lipid bilayer itself must be an essential, or even a determining, contribution to the interaction between almost all proteins known to generate local membrane curvatures[1,2,5,6]. Provided that this attraction is not opposed by an equally strong repulsion coming, say, from the electric charges in the proteins, it could be the driving force for membrane shaping in real biological conditions.

Verification of the results obtained by Reynwar and colleagues[3] is needed through other simulation models and an understanding of the forces' exact physical origin. Assuming that verification is attained, membrane-mediated attractive forces will join the physical arsenal that is used to probe the very biological business of membrane proteins. ■

Michael M. Kozlov is in the Department of Physiology and Pharmacology, Sackler Faculty of Medicine, Tel Aviv University, 69978 Tel Aviv, Israel.
e-mail: michk@post.tau.ac.il

1. McMahon, H. T. & Gallop, J. L. *Nature* **438,** 590–596 (2005).
2. Zimmerberg, J. & Kozlov, M. M. *Nature Rev. Mol. Cell Biol.* **7,** 9–19 (2006).

3. Reynwar, B. J. *et al. Nature* **447,** 461–464 (2007).
4. Luini, A. *et al. Curr. Opin. Cell Biol.* **17,** 353–361 (2005).
5. Itoh, T. & De Camilli P. *Biochim. Biophys. Acta* **1761,** 897–912 (2006);
6. Shibata, Y. *et al. Cell* **126,** 435–439 (2006).
7. Bruinsma, R. & Pincus, P. *Curr. Opin. Solid State Mater. Sci.* **1,** 401–406 (1996).

MOLECULAR MEDICINE

MicroRNAs and the tell-tale heart

Kenneth R. Chien

MicroRNAs are natural, single-stranded, small RNA molecules thought to control gene expression. Four studies indicate that specific microRNA sequences can regulate heart function in development and disease.

Understanding how complex physiological processes are coordinately controlled at the molecular level *in vivo* is one of the cornerstones of research in modern translational (bench-to-bedside) medicine. Using the heart as a model physiological system, four research teams[1–4] now establish a role for microRNAs (miRNAs or miRs) in the regulation of *in vivo* cardiac function: specifically, the conductance of electrical signals, heart muscle contraction, and heart growth and morphogenesis. These studies also have implications for understanding pathways involved in heart disease, and point to potential opportunities and challenges in manipulating miRNAs as therapeutic targets in general.

By creating mice that are deficient in a muscle-specific miRNA, miR-1-2, Zhao *et al.*[1] uncover a role for miRNAs in one of the commonest forms of congenital heart disease — ventricular septal defects, characterized by a hole in the wall between the left and right ventricles of the heart. Such mice also show marked abnormalities in cardiac conduction, and suffer from hyperplasia (an increase in the number of cardiac muscle cells), which leads to heart enlargement[1]. In a separate study, Yang *et al.*[2] report that overexpressing miR-1 in the hearts of adult mice has a prominent effect on the development of cardiac arrhythmia — irregular electrical activity in the heart, which can be life threatening. This indicates that a tight regulation of miR-1 levels is crucial for the maintenance of normal cardiac conduction. Both studies identify miR-1 targets that may, at least in part, account for the manifestation of the associated diseases. These targets are a gap-junction protein that functions in the electrical communication between cardiac cells[2], and a cardiac transcription factor, Irx5, which regulates the formation and maturation of the electrical conduction system in the heart[1].

In the third study, Care *et al.*[3] focus on the muscle-specific miR-133, and find that it is a negative regulator of cardiac hypertrophy — an essential adaptive physiological response to mechanical and hormonal stress — whereby cardiac cells grow in size. Moreover, van Rooij and colleagues[4] find that miR-208 — a heart-specific miRNA that is transcribed from a sequence embedded in the non-coding region of the gene for the heavy chain of the α-myosin protein — also has a modulatory role in controlling the hypertrophic response.

In the mouse heart, the main function of miR-208 seems to be mediating the switch from expression of the heavy chain of α-myosin to that of β-myosin during stress or thyroid-hormone-induced cardiac growth[4]. This effect might be linked to repression of the thyroid hormone co-regulator THRAP1, which can function as both a positive and a negative regulator of transcription[4].

Taken together, these studies provide clear evidence that a subset of miRNAs can modulate a diverse spectrum of cardiovascular functions *in vivo* (Fig. 1). The findings have implications for the main milestones along the highway of translational medicine — developmental, physiological and clinical. Although these observations indicate that miRNAs affect cardiac muscle, their effects are probably exerted at the level of heart-muscle-cell lineages, such as embryonic and adult ventricular muscle cells, ventricular septal muscle cells and the muscle cells of the conduction system.

Do these findings reflect a role for miRNAs in the control of specific cell-fate decisions during the course of heart-cell-lineage formation or maturation? Mice that suffer a global loss of miRNAs after the conditional ablation of the gene that encodes Dicer — a protein involved in the processing of miRNAs — in early heart progenitor cells show a range of cardiac defects[1], suggesting that miRNAs control specific cell-fate decisions. Recently, a hierarchy of embryonic and postnatal cardiovascular progenitor cells, marked by the expression of the transcription factor islet-1, has been discovered that can differentiate into cardiac smooth muscle or endothelial-cell lineages[5,6]. This implies that the diversification of heart-cell lineages is controlled in a similar way to the differentiation of blood stem cells. Moreover, other cardiovascular progenitor cells have been described[7–10], and the use of these cells, along with well-characterized assays using mouse embryonic stem cells[5,7,8] and mouse embryo model systems[5,7,9], suggests that assignment of these miRNAs to specific locations in the fate map of cardiovascular cell lineages should be forthcoming shortly. Thus, the identification of the targets of these miRNAs might eventually lead to new markers for specific cardiovascular progenitor cells and their differentiated progeny.

From a physiological perspective, these studies[2–4] indicate that an array of complex cardiovascular endpoints can be affected by either gain or loss of function of a diverse group of miRNAs, even after birth. Given this great sensitivity, and their broad effects on a wide range of genes specifically expressed in the heart, it is possible that slight variations in the miRNA levels have profound effects

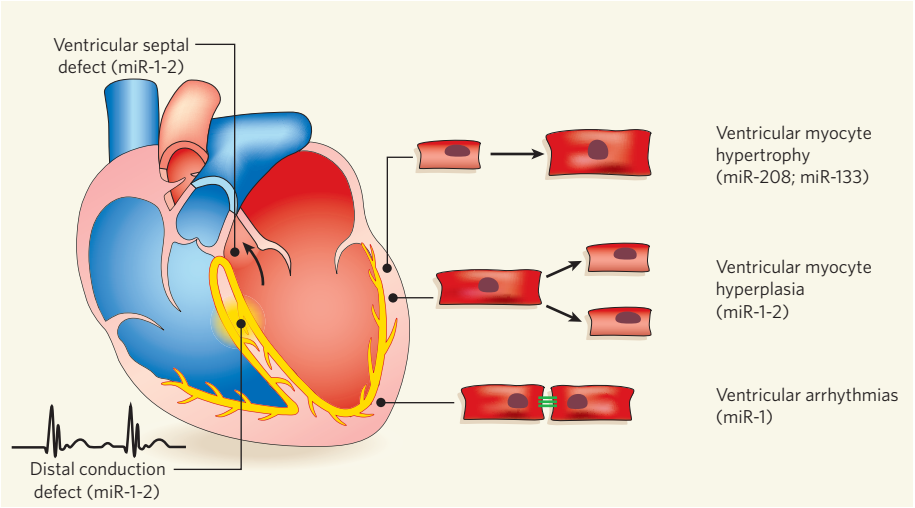


Figure 1 | MicroRNAs and heart function. Four papers[1–4] implicate certain microRNAs (miRs) in regulating diverse processes linked to health and disease in the heart. Ventricular septal defects involve a hole in the wall between the right and left ventricles of the heart. Distal conduction defect relates to disease in the muscle cells of the conduction system and results in irregular broadening of the electrical conduction pattern. Ventricular myocyte hypertrophy indicates an enlargement of heart muscle cells, whereas ventricular myocyte hyperplasia describes an increase in the number of these cells. In ventricular arrhythmias, the electrical impulse of the heartbeat is irregular.

on heart health and disease. Genome-wide association studies continue to uncover new genetic variants involved in the control of cardiac electrical conduction[11,12], as well as overt diabetes and related coronary heart disease[13–15]. So it will be interesting to determine whether miRNAs, such as those described in these studies[1–4], might correlate with the traits discovered through genome-wide association studies that are implicated in common forms of heart disease. Furthermore, it is possible that naturally occurring variations in miRNAs themselves, or in their molecular control, would be directly linked to the genetics of complex cardiovascular traits in human populations.

These findings[1–4] reveal a level of molecular control of heart physiology that is beyond the well-accepted regulatory role of signalling pathways — mediated by kinases, transcription-factor complexes and alterations in tissue-specific structural proteins — in the heart. These results are beginning to reshape our view of how a spectrum of complex cellular physiological functions can be coordinately regulated by a handful of tissue-specific miRNAs.

Converting fundamental observations such as these[1–4] into biologically targeted cardiovascular therapy is a central tenet in translational medicine. The challenge will be to control off-target effects, to design approaches for achieving a long duration of effectiveness, which is necessary for the treatment of chronic heart disease, and to carefully titrate the expression level of a given miRNA so that it affects a selective, disease-related endpoint. In doing so, perhaps we will be entering a new era of 'pre-translational' medicine. ■

Kenneth R. Chien is in the MGH Cardiovascular Research Center, Department of Cell Biology, Harvard Medical School and the Harvard Stem Cell Institute, Boston, Massachusetts 02114, USA.
e-mail: kchien@partners.org

1. Zhao, Y. *et al. Cell* **129,** 303–317 (2007).
2. Yang, B. *et al. Nature Med.* **13,** 486–491 (2007).
3. Care, A. *et al. Nature Med.* **13,** 613–618 (2007).
4. van Rooij, E. *et al. Science* **316,** 575–579 (2007).
5. Moretti, A. *et al. Cell* **127,** 1151–1165 (2006).
6. Laugwitz, K. L. *et al. Nature* **433,** 647–653 (2005).
7. Wu, S. M. *et al. Cell* **127,** 1137–1150 (2006).
8. Kattman, S. J., Huber, T. L. & Keller, G. M. *Dev. Cell* **11,** 723–732 (2006).
9. Prall, O. W. *et al. Cell* **128,** 947–959 (2007).
10. Smart, N. *et al. Nature* **445,** 177–182 (2007).
11. Newton-Cheh, C. & Shah, R. *Curr. Opin. Genet. Dev.* doi:10.1016/j.gde.2007.04.010 (2007).
12. Arking, D. E. *et al. Nature Genet.* **38,** 644–651 (2006).
13. Saxena, R. *et al. Science* doi:10.1126/science.1142358 (2007).
14. McPherson, R. *et al. Science* doi:10.1126/science.1142447 (2007).
15. Helgadottir, A. *et al. Science* doi:10.1126/science.1142842 (2007).

CONDENSED-MATTER PHYSICS

A superfluid is born

Henk T. C. Stoof

For most of its existence, a superfluid droplet leads an essentially innocuous, classical life. But intense scrutiny reveals that the birth of such droplets is a turbulent and unpredictable quantum affair.

Phase transitions occur everywhere in nature. They can happen on a grand, cosmic scale: our Universe is thought to have gone through several phase transitions in the first second after the Big Bang. Equally, they can be mundane, small-scale events: the condensation of water vapour into droplets on a cold surface is an everyday example. The ubiquity of phase transitions makes understanding exactly how they develop over time of essential interest. In two beautiful experiments, Hugbart *et al.*[1], writing in *Physical Review A*, and Ritter *et al.*[2], writing in *Physical Review Letters*, take an intimate look at how a phase transition similar to, but more exotic than, the gas–liquid transition occurs — the birth of a superfluid droplet.

This particular phase transition occurs when a trapped atomic gas is cooled almost to the absolute zero of temperature, forming what is known as a Bose–Einstein condensate. This condensate droplet is liquid-like. But it is also fundamentally different, in that a large object can move through it without friction. Thanks to a combination of experimental and theoretical work[3–7], the growth of a superfluid droplet after birth from the surrounding atomic gas is already rather well understood. Looked at at the microscopic level, the growth occurs as a result of collisions between two atoms in which one atom is stimulated to end up in the condensate. A similar thing happens in a laser, where the particles of the laser medium are excited so that they preferentially emit a photon into the laser beam. Because of the stimulated nature of the condensate growth, and because the collisions occur at a well-defined position in the gas, the process can be understood essentially classically. Quantum mechanics does not have a significant role.

The actual birth of a Bose–Einstein condensate is expected to be a different kettle of fish. Indeed, the famous Kibble mechanism[8] predicts that, at the moment of its birth, a superfluid droplet contains a highly dynamic tangle of tiny quantum tornadoes[9]. For a trapped atomic gas, however, this mechanism is unlikely[10]. For one thing, the smallest possible size of the superfluid droplet, determined by Heisenberg's uncertainty principle, is tiny. The formation of a quantum-tornado tangle in such small droplets would be energetically very costly. For another thing, the usual evapora-

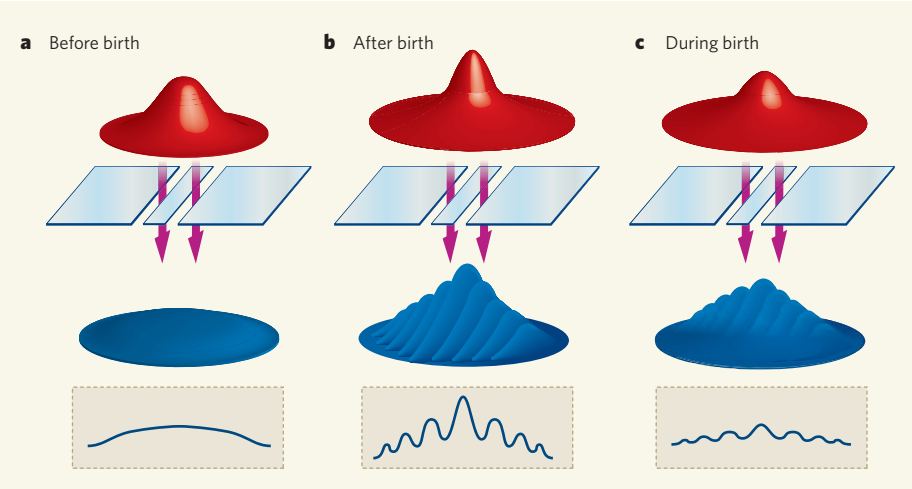


Figure 1 | Stages in superfluid development. In this experiment, essentially that performed by Ritter *et al.*[2], a cloud of ultracold atoms falls through a double slit onto a detector. **a**, Before the birth of the superfluid droplet, the atomic cloud has a normal (gaussian) density profile (red). As the atoms of the gas are uncorrelated, the density profile at a detector on the other side of the double slit is merely a spread-out distribution, with no interference pattern. **b**, In a Bose–Einstein condensate (bump in initial profile), all atoms are correlated in phase. Atoms coming through the two slits have a precise phase difference, leading to a predictable pattern of interference on the detector. This behaviour can still be called classical, because it is in essence identical to the classical interference of waves with a definite phase relationship. **c**, During birth, however, the superfluid droplet is a fuzzy, quantum object. This is seen as a reduced visibility of the interference pattern and, if the Kibble mechanism were at work, even as unpredictable shifts in the pattern when the experiment is repeated.

tive cooling method that is used to create the superfluid droplet affects only the most energetic atoms in the gas. A more likely model, therefore, is that at the birth of a superfluid droplet, its energetically lowest-lying collective vibrations are excited by quantum 'noise' in the surrounding atomic gas. The result is a 'fuzzy' quantum object, whose strongly oscillating nature initially hinders growth. After these vibrations have largely damped down, the droplet can grow further by the classical, collisional process already mentioned.

So how can we observe the quantum nature of the nascent superfluid droplet? This requires a measurement that is sensitive to non-local properties spread over a large part of the Bose–Einstein condensate. A particularly dexterous possibility is to perform an interference experiment in which the atomic cloud falls through a double slit (Fig. 1). Ritter *et al.*[2] in essence perform this experiment, with a double slit created by microwave fields that preferentially extract atoms from the condensate in two different places, and an interference pattern analysed by looking at the number of detected atoms as a function of time.

Hugbart *et al.*[1] implement another possibility, using an ingenious spectroscopic method to measure the number of atoms in the Bose–Einstein condensate that have a certain velocity. Notwithstanding the quite different methods, both experiments clearly show the fuzzy, vibrating quantum nature of the superfluid droplet at birth. The main difference between the findings of the two experiments lies in the precise nature of the vibrations that are excited. Hugbart *et al.* observe quadrupole-shaped oscillations of the droplet, whereas Ritter *et al.* observe subtle fluctuations in superfluid flow that hardly affect the shape of the droplet. This difference is possibly related to the much more elongated shape of the superfluid droplet in Hugbart and colleagues' experiment, although more work is required to clarify the reasons for this qualitative difference.

These experiments[1,2] allow us a first glimpse of the birth and initial growth of a Bose–Einstein condensate. That is already leading to new qualitative insights into this fundamental process, but the experimental results cry out for the detailed, quantitative comparison with theory that will allow a precise understanding of the process. A further attractive option is to study the birth of the superfluid droplet in the presence of an optical lattice — a web of laser beams that creates a regular array of sites to which the atoms of the ultracold gas are attracted. For a sufficiently attractive optical lattice, the Bose–Einstein condensate would have to form not from a normal gas, but from yet another phase of matter, a Mott insulator, in which every site of the optical lattices is filled with exactly one atom. Ritter and colleagues also have experience with creating these states, so such an experiment might not be too difficult for them to carry out. ■

Henk T. C. Stoof is at the Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, the Netherlands.
e-mail: h.t.c.stoof@phys.uu.nl

1. Hugbart, M. *et al. Phys. Rev. A* **75,** 011602 (2007).
2. Ritter, S. *et al. Phys. Rev. Lett.* **98,** 090402 (2007).
3. Miesner, H.-J. *et al. Science* **279,** 1005–1007 (1998).
4. Davis, M. J., Gardiner, C. W. & Ballagh, R. J. *Phys. Rev. A* **62,** 063608 (2000).
5. Bijlsma, M. J., Zaremba, E. & Stoof, H. T. C. *Phys. Rev. A* **62,** 063609 (2000).
6. Köhl, M. *et al. Phys. Rev. Lett.* **88,** 080402 (2002).
7. Shvarchuck, I. *et al. Phys. Rev. Lett.* **89,** 270404 (2002).
8. Kibble, T. W. B. *J. Phys. A* **9,** 1387–1398 (1976).
9. Svistunov, B. *Phys. Lett. A* **287,** 169–174 (2001).
10. Stoof, H. T. C. *J. Low Temp. Phys.* **114,** 11–108 (1999).

MOLECULAR BIOLOGY

RNA in control

Benjamin J. Blencowe and May Khanna

In bacteria, some messenger RNAs can sense the need for their protein product and accordingly regulate expression of their own genes. A similar type of RNA regulation has now been revealed in higher organisms.

The functional capacity of RNA, beyond its role in protein synthesis, frequently amazes, as examples of RNA-mediated gene regulation are continuously emerging. One type of such regulation found in bacteria involves RNA structures called riboswitches. These are sequences of nucleotide bases in messenger RNAs that contain structural domains called aptamers. Aptamers act as sensors by binding to a specific small-molecule building-block, or metabolite. The protein product of the riboswitch mRNA is often involved in the biosynthesis or transport of the same metabolite[1]. On binding to a metabolite, aptamers undergo a conformational change that alters the mRNA's access to the machinery required for either its transcription from a gene or its translation into a protein. Thus, riboswitches regulate the intracellular levels of bacterial metabolites. However, it was not known whether they have similar functions in eukaryotes (fungi, plants and animals). On page 497 of this issue, Cheah and colleagues[2] reveal a mechanism by which riboswitches regulate the expression of genes involved in vitamin B_1 biosynthesis in the fungal species *Neurospora crassa*.

50 YEARS AGO

Completion of construction on the Dounreay fast reactor is expected towards the end of the year... One of the main objects for the Dounreay reactor will be to develop fuel elements capable of burning fissile atoms rapidly, and able to withstand high heat ratings with high outlet temperatures for the coolant. The use of plutonium as a fuel will be investigated, to illustrate the economics of a system based on the breeding of plutonium from natural or depleted uranium. The economics of the chemical processes required to handle highly active irradiated fuel will form part of this study, since there is no requirement for separation of 'poisons' in the shape of fission products with high capture cross-section for neutrons, but rather a means of re-forming fuel which may have suffered extensive mechanical damage from the fission process.
From *Nature* 25 May 1957.

100 YEARS AGO

The Khasis are a tribe inhabiting the Khasi and Jaintia Hills in the Indian province of (as it is now called) Eastern Bengal and Assam. They are surrounded on all sides by alien peoples, Tibeto-Burman and Aryan, and are believed to be a survival of a primitive Austro-Asiatic race that once occupied the whole of eastern India until they were conquered and dispossessed in prehistoric times by an invasion of Tibeto-Burmans. The tribal constitution is strongly matriarchal. Inheritance is through the female line, the youngest daughter being the chief heir of her mother; ancestral property can only be owned by women, and the only property which a man can possess is that which is self-acquired. The chief deities are all female. So is the sun, while the moon is represented as a man, and in grammar and vocabulary the feminine element is much more prominent than the masculine.
From *Nature* 23 May 1907.

50 & 100 YEARS AGO

NATURE|Vol 447|24 May 2007

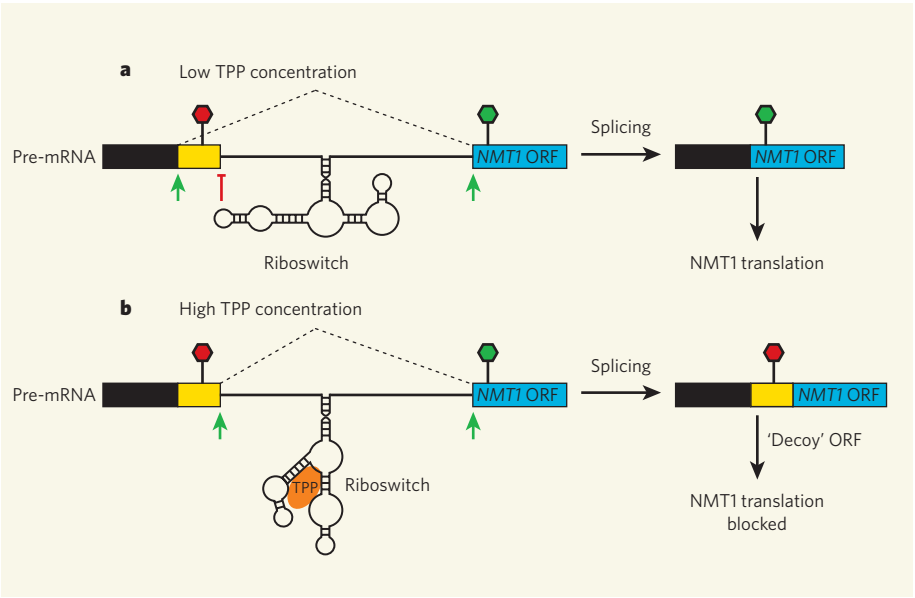


Figure 1 | Regulation of gene expression by riboswitches in *Neurospora crassa*. Cheah *et al.*[2] show that expression of the *NMT1* gene is regulated at the level of pre-mRNA alternative splicing by a riboswitch that binds to thiamine pyrophosphate (TPP). **a**, At low concentrations of TPP, the TPP-binding (aptamer) region of the riboswitch base-pairs with sequences surrounding a splice site (red blocking line) in a nearby non-coding sequence, and prevents its selection by the splicing machinery. A distal splice site (green arrow) is selected, however, resulting in the generation of a shorter *NMT1* mRNA with a coding sequence, or open reading frame (ORF), that translates into a functional *NMT1*-encoded protein (green signal). **b**, At high TPP levels, the aptamer undergoes a conformational rearrangement so that the region that was previously bound to the nearby splice site is now used to bind to TPP. This and other conformational changes (not shown) generate a longer mRNA splice variant that contains short, 'decoy' ORFs (red signal), preventing functional *NMT1* expression.

Unlike coding sequences in bacterial genes, which typically are continuous stretches of DNA, eukaryotic genes contain coding regions (exons) that are separated by non-coding sequences (introns). These genes are initially transcribed into precursor (pre)-mRNAs, which undergo a series of processing steps, including the removal of introns and splicing together of exons, before mature mRNA is generated. A complex macromolecular machinery in the nucleus of eukaryotic cells is responsible for pre-mRNA splicing[3].

Alternative splicing occurs when the boundaries between exons and introns, called the splice sites, are differentially selected by the splicing machinery to generate two or more different mRNAs from the same pre-mRNA. Alternative splicing is a remarkably efficient mechanism for a cell to increase the structural and functional diversity of its proteins, and it plays many roles in gene regulation[4–6]. In almost every known case, the regulation of alternative splicing is achieved through an interplay between protein factors and pre-mRNA sequences at or near the splice sites. The fungal riboswitches identified by Cheah *et al.* break this mechanistic mould, as their action in dictating the choice of splice sites can be explained entirely by a metabolite-induced change in RNA folding, without the direct involvement of proteins.

An initial clue suggesting the possibility of riboswitch-mediated regulation of splicing in eukaryotes came from the observation that, in fungal and plant species[7,8], evolutionarily conserved blocks of sequences, which resemble bacterial aptamers and bind to thiamine pyrophosphate (TPP), are present in the introns of genes involved in thiamine biosynthesis. TPP is a derivative of thiamine, commonly known as vitamin B_1. It was subsequently shown that in the fungus *Aspergillus oryzae*, deletion of a TPP aptamer from the intron of the *thiA* gene (involved in thiamine biosynthesis) prevents this gene from responding when thiamine is added to the growth medium[9].

Cheah and colleagues[2] now show that the addition of thiamine to the growth medium of *N. crassa* results in marked shifts in the alternative splicing patterns of mRNAs for three genes that all contain TPP aptamers. Two of these genes, *NMT1* and *THI4*, are involved in thiamine metabolism, whereas the third gene encodes a protein of unknown function[2]. For *NMT1* and *THI4*, with high concentrations of thiamine, the levels of mRNA splice variants lacking functional protein-coding sequences were increased relative to the levels of splice variants corresponding to functional mRNAs. How do these shifts in alternative splicing patterns occur?

As mentioned earlier, bacterial riboswitches function by undergoing metabolite-induced structural changes, which alter the access of the mRNAs to their transcriptional or translational machineries. A related mechanism involves the TPP aptamer in the *NMT1* mRNA. At low TPP levels, the region surrounding the aptamer is folded such that it favours selection of a distal splice site. The resulting shorter mRNA contains a coding sequence, also called an open reading frame (ORF), for a functional *NMT1*-encoded protein (Fig. 1a). In this conformation, bases in a section of the aptamer pair up with bases overlapping a competing splice site on a nearby non-coding sequence in the pre-mRNA, thereby preventing it from being selected to produce non-functional mRNA.

At high concentrations, TPP binds tightly to the aptamer and causes a conformational rearrangement that prevents it from base-pairing with the nearby splice site. Consequently, this competing splice site is selected by the splicing machinery, and a longer mRNA splice variant is produced, which contains short, 'decoy' ORFs that prevent the expression of the main *NMT1* ORF (Fig. 1b). A feature of this switch is that, although aptamer sequences can bind to both TPP and the competing splice site, the two events are mutually exclusive.

In addition to its unique mechanism of action, the splicing-regulatory TPP riboswitch in *N. crassa* has another intriguing feature: its function resembles more complex strategies that use alternative splicing to coordinate gene activities in higher eukaryotes. For example, the finding that a single metabolite can alter the splicing patterns of at least two genes that operate in the same biochemical pathway is reminiscent of other observations indicating that alternative splicing can be regulated in a coordinated manner to control functionally related genes[6]. It seems plausible that splicing-regulatory riboswitches represent a system that has evolved to coordinately regulate multiple genes in the same biochemical pathway using feedback and, in some cases, feed-forward mechanisms. Presumably, the rapid kinetics and energy-saving advantages afforded by bypassing protein-mediated regulation explain why riboswitch aptamers have persisted during evolution and function at many levels of regulation of gene expression.

This raises a question: to what extent do riboswitches regulate alternative splicing, or other steps in gene expression, in eukaryotes? The answer might be, quite often. In the fungus *Aspergillus nidulans*[10], a splicing-regulatory riboswitch that binds to the amino acid L-arginine was recently discovered in the non-coding region of a gene that encodes arginase — an enzyme required for using arginine as a source of energy. However, a natural role for small-molecule aptamers in regulating splicing has not been found in animal cells, although an artificial riboswitch has been engineered to control splicing in cultured mammalian cells[11].

Given the discovery of riboswitch aptamers that regulate alternative splicing in fungi[2], and the possible existence of equivalent mechanisms in plants[12], it is reasonable to expect that animals might utilize related strategies

for regulating gene expression. With advances in computational strategies for locating conserved RNA folds in sequence databases, high-throughput methods for monitoring alternative splicing and other steps in gene expression, and prior knowledge of the function of genes involved in small-molecule metabolism, finding other examples of such regulatory modules in eukaryotes seems possible. In any case, we can be almost certain that new forms of RNA-based regulation will continue to emerge and amaze. ■

Benjamin J. Blencowe and May Khanna are in the Banting and Best Department of Medical Research, the Department of Molecular and Medical Genetics and the Centre for Cellular and Biomolecular Research, Donnelly CCBR Building, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada.
e-mail: b.blencowe@utoronto.ca

1. Tucker, B. J. & Breaker, R. R. *Curr. Opin. Struct. Biol.* **15,** 342–348 (2005).
2. Cheah, M. T., Wachter, A., Sudarsan, N. & Breaker, R. R. *Nature* **447,** 497–500 (2007).
3. Jurica, M. S. & Moore, M. J. *Mol. Cell* **12,** 5–14 (2003).
4. Graveley, B. R. *Trends Genet.* **17,** 100–107 (2001).
5. Matlin, A. J. *et al. Nature Rev. Mol. Cell Biol.* **6,** 386–398 (2005).
6. Blencowe, B. J. *Cell* **126,** 37–47 (2006).
7. Sudarsan, N. *et al. RNA* **9,** 644–647 (2003).
8. Galagan, J. E. *et al. Nature* **438,** 1105–1115 (2005).
9. Kubodera, T. *et al. FEBS Lett.* **555,** 516–520 (2003).
10. Borsuk, P. *et al. Biol. Chem.* **388,** 135–144 (2007).
11. Kim, D. S. *et al. RNA* **11,** 1667–1677 (2005).
12. Thore, S. *et al. Science* **312,** 1208–1211 (2006).

SUPERNOVAE

Answers and questions

David Branch and Ken'ichi Nomoto

Do we understand the violent and cosmologically significant stellar explosions known as type-Ia supernovae? Yes and no, as astronomers participating in a conference in California agreed.

In mid-March, more than 100 astronomers converged on the Kavli Institute for Theoretical Physics in Santa Barbara, California, for an international conference* on so-called type-Ia supernovae (SNe Ia). Understanding these stellar explosions has a high priority: measurements of their brightness in the late 1990s revealed the existence of a mysterious 'dark energy' permeating space and accelerating the Universe's expansion. This conference was not primarily about exploiting SNe Ia for cosmology, but about assessing our current state of knowledge of where they come from, what exactly their stellar progenitors are, how they work, and how they explode.

A good idea lasts

In 1960, Fred Hoyle and William Fowler[1] concluded that SNe Ia are the result of thermonuclear instability following the ignition of nuclear fuel in 'electron-degenerate' matter. Such matter is formed when a star contracts and the electrons of its matter are compressed to fill every energy level available to them by the quantum-mechanical Pauli exclusion principle. Since then, astronomers have fleshed out the idea. Unlike other supernovae — types Ib, Ic and II, collectively known as core-collapse supernovae and produced only by short-lived, massive stars — SNe Ia are seen in both young and old stellar populations. They are even found in elliptical galaxies, meaning that some of them are produced by long-lived, low-mass stars found in these galaxies.

Most low-mass stars end their lives as electron-degenerate carbon–oxygen white dwarfs, without exploding. A more dramatic fate comes if the white dwarf accretes non-degenerate matter from a companion in a binary system (the single-degenerate scenario) or merges with

Paths to Exploding Stars: Accretion and Explosion, Santa Barbara, California, 19–23 March 2007; http://online.kitp.ucsb.edu/online/snovae_c07

HYDROLOGY

Flood of data

P. BATES

If you need more precise measurements of natural events on Earth's surface, get into space. Researchers studying glaciers and earthquakes have for some time followed this principle, exploiting the power of satellite interferometric imaging to map surface displacements down to the centimetre scale. Doug Alsdorf and his colleagues have taken the same approach in their investigations of the periodic floods that occur in the Amazon basin (D. Alsdorf *et al. Geophys. Res. Lett.* **34,** doi:10.1029/2007GL029447; 2007).

The Amazon river has an intimate relationship with its vast floodplain, with an estimated 25% of its average annual discharge flowing and ebbing across it. But very little is known about the behaviour of these floods: not least, gauges of water level are placed only along the main channels, and then only sparsely. There are technical difficulties in taking interferometric measurements of water surfaces with satellite-borne synthetic-aperture radar. But flooded vegetation (pictured) does reflect an adequate signal.

Using data provided by instruments aboard the Japanese Earth Resources Satellite, Alsdorf *et al.* have been able to map the spatial and temporal complexity of floodplain inundation. Their study of floods from three different years takes in an area of the central Amazon basin that includes flows from the Purus river, as well as the Amazon itself.

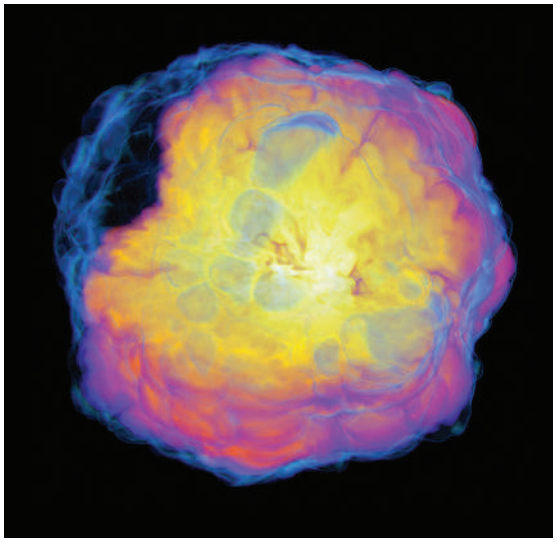
Water levels in the floods, it turns out, do not take on the pattern that might be expected from a simple correspondence with the levels in the main channel of the river. Rather, there is a complicated interplay in which flow paths and water levels are influenced not only by the main channel and floodplain topography, but also by local and far-reaching hydraulic factors created by the flood itself.

These are proof-of-principle findings, with a practical edge. Modelling of floods is bedevilled by a lack of relevant measurements to test them. Satellite data can help redress that lack, with the ultimate aim of guiding engineering or other solutions to the inundation of areas inhabited by human populations. Furthermore, periodic flooding, and the associated delivery of sediments and nutrients, is a natural feature of wetland ecosystems not only in the Amazon but throughout the world. Some wetlands are under threat and, in some, restoration projects are in hand. Clarification of the relevant networks of water flow in different circumstances would offer another approach to ensuring the long-term success of such projects.

Tim Lincoln

another white dwarf (the double-degenerate scenario) to form a configuration that approaches or exceeds the limiting 'Chandrasekhar' mass for a white dwarf, which amounts to 1.4 solar masses. Then, ignition of nuclear fuel near the centre of the body, and the outward propagation of a nuclear burning front, can release enough energy to explode the star within a second, leaving nothing behind (except, in the single-degenerate scheme, a toasted companion star).

The theoretically calculated composition and density structure of such explosions are expected to be broadly consistent with the observed spectra and light curves (of brightness against time) of SNe Ia. Indeed, increasingly detailed calculations of spectra and light curves for suitably parametrized numerical models of exploding white dwarfs are providing impressive agreement with observations (D. Kasen, Johns Hopkins Univ.; S. Woosley, UCSC)[2]. A white dwarf that accretes matter until it explodes still seems to be the best idea going.



F. K. ROEPKE/W. HILLEBRANDT/MAX-PLANCK-INST. ASTROPHYS., GARCHING

Figure 1 | Burn-out. A three-dimensional calculation of a pure deflagration explosion of a white dwarf. Blue areas represent high-velocity, low-density, unburnt carbon and oxygen; yellow represents low-velocity, high-density matter that has burnt to iron-group elements. In the central part of the image, the blue has been removed in order to make the interior yellow visible.

Hunt for white dwarfs

A large galaxy such as ours produces SNe Ia about once a century. At any time, therefore, it should contain numerous accreting white dwarfs slowly approaching catastrophe. A few single-degenerate systems in which an accreting white dwarf seems to be near the Chandrasekar mass have been identified (I. Hachisu, Univ. Tokyo; J. Sokoloski, Columbia Univ.)[3,4], but in general these systems seem to be in short supply.

A favourite class of single-degenerate progenitor has been 'super-soft-X-ray sources', many of which are white dwarfs that accrete matter from non-degenerate companion stars. These white dwarfs gain mass by undergoing steady nuclear burning near their surfaces, thus emitting low-energy 'soft' X-rays. But recent surveys of such sources in nearby galaxies with NASA's Chandra X-ray Observatory have detected far too few of them to account for the observed rate of SNe Ia (R. Di Stefano, Harvard Univ.). It might be that recurrent, weaker nova eruptions and the absorption of X-rays by winds from the hot white dwarfs substantially reduce the observable duration of soft-X-ray emission.

Because white dwarfs are very dim, double-degenerate systems can be found only in a small, nearby volume of space. Most of the systems found so far have a total mass that is less than the Chandrasekhar limit, but one system whose total mass seems to be beyond it has been found (R. Napiwotzki, Univ. Hertfordshire)[5]. Double-degenerate systems might help to account for SNe Ia in very old populations such as elliptical galaxies. But whether the merging stars will really explode, rather than collapse to form neutron stars, is unclear (P. Podsiadlowski, Oxford Univ.). X-rays from the products formed by the merger, from before the supernova occurs, are not seen in sufficient numbers (R. Webbink, UIUC).

Mystery of exploding dwarfs

For decades we have known that, if nuclear ignition initially produces an outgoing detonation front that propagates supersonically, the entire white dwarf will be burnt to iron-group isotopes. This is inconsistent with the signatures of the intermediate-mass elements magnesium, silicon, sulphur and calcium in SNe Ia spectra. Instead, ignition must initially produce a 'deflagration', a form of combustion that propagates subsonically.

One-dimensional, spherically symmetrical calculations of nuclear hydrodynamics can produce the composition and density structures needed to match light curves and spectra. Computationally intensive calculations in three dimensions, however, reveal that pure deflagration explosions are relatively weak (W. Hillebrandt and F. Röpke, MPA Garching) (Fig. 1). A 'deflagration–detonation' transition similar to that which occurs in the cylinders of a car might provide more energy, but whether a nuclear equivalent could occur in the unconfined ejecta of a supernova is unclear (V. Gamezo, US Naval Research Lab.; L. Dursi, Univ. Toronto)[6,7].

Another possibility — gravitationally confined detonation — invokes a buoyant bubble of burnt fuel that bursts through the surface of the white dwarf and drives a 'flood' across the surface that converges at the antipode, causing a detonation (T. Plewa and D. Lamb, Univ. Chicago)[8,9]. This and all other explosion models involve thorny issues of nuclear-combustion physics; which of the scenarios are viable remains a subject of lively debate[10].

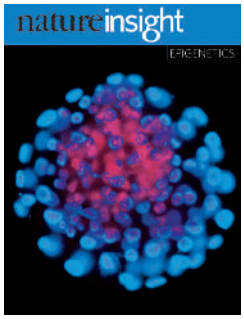
Saving grace

The consensus in Santa Barbara remained that SNe Ia are produced by accreting or merging white dwarfs. But as observational advances occur and more complex explosion models are calculated, the questions of where the progenitors are and how they explode seem to become only more perplexing. Beyond the issue of how nature makes the impressively homogeneous 'normal' SNe Ia, questions persist about what causes the moderate diversity among them, as well as what causes the minority that are strikingly peculiar in various ways (A. Filippenko, UC Berkeley). These include a homogeneous class of events of unusually low energy (M. Phillips, Las Campanas Observ.)[11–13], and one event the mass of whose ejecta seems to have been substantially above the Chandrasekhar limit (D. Howell, Univ. Toronto)[14].

In the face of these perplexing features, a saving grace is that the use of SNe Ia luminosity as empirical distance indicators for cosmology — the foundation on which the first observation of the Universe's accelerating expanion was built — still seems to be valid. Mild differences in the intrinsic luminosity of normal SNe Ia can, to a first-order approximation, continue to be allowed for through an observed relation between peak luminosity and light-curve width. Extreme outliers such as the 'super-Chandra' event[14] can be excluded from cosmological samples. Nevertheless, the use of SNe Ia for cosmology will be on firmer ground when our understanding of the phenomenon itself improves. Observations of SNe Ia, the search for their progenitors, and efforts to calculate properly the propagation of their nuclear flame are set to continue. ■

David Branch is in the Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma 73019, USA. Ken'ichi Nomoto is in the Department of Astronomy and the Research Center for the Early Universe, School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan.
e-mails: branch@nhn.ou.edu; nomoto@astron.s.u-tokyo.ac.jp

1. Hoyle, F. & Fowler, W. A. *Astrophys. J.* **132,** 565–590 (1960).
2. Kasen, D. & Woosley, S. E. *Astrophys. J.* **656,** 661–665 (2007).
3. Hachisu, I. *et al. Astrophys. J. Lett.* **659,** 153–156 (2007).
4. Sokoloski, J. *et al. Nature* **442,** 276–278 (2006).
5. Geier, S. *et al. Astron. Astrophys.* **464,** 299–307 (2007).
6. Gamezo, V., Khokhlov, N. & Oran, E. S. *Astrophys. J.* **623,** 337–346 (2005).
7. Zingale, M. & Dursi, L. J. *Astrophys. J.* **656,** 333–346 (2007).
8. Plewa, T. *Astrophys. J.* **657,** 942–960 (2007).
9. Jordan, G. C. IV *et al.* preprint available at www.arxiv.org/astro-ph/0703573 (2007).
10. Roepke, F. K. *et al. Astrophys. J.* **660,** 1344–1356 (2007).
11. Li, W. *et al. Publ. Astron. Soc. Pacif.* **115,** 453–473 (2003).
12. Jha, S. *et al. Astron. J.* **132,** 189–196 (2006).
13. Phillips, M. M. *et al. Publ. Astron. Soc. Pacif.* **119,** 360–387 (2006).
14. Howell, D. A. *et al. Nature* **443,** 308–310 (2006).



Cover illustration
Bovine blastocyst cells that give rise to different tissues show differential DNA methylation (pink). (Courtesy of F. Santos, W. Dean and W. Reik.)

EPIGENETICS

Epigenetics is typically defined as the study of heritable changes in gene expression that are not due to changes in DNA sequence. Diverse biological properties can be affected by epigenetic mechanisms: for example, the morphology of flowers and eye colour in fruitflies.

Epigenetic changes are crucial for the development and differentiation of the various cell types in an organism, as well as for normal cellular processes such as X-chromosome inactivation in female mammals and silencing of mating-type loci in yeast. However, epigenetic states can become disrupted by environmental influences or during ageing, and the importance of epigenetic changes in the development of cancer and other diseases is increasingly being appreciated.

With the elucidation of a molecular basis for epigenetics, the field has flourished, even though the mechanisms of heritability are often obscure. Epigenetic processes can involve chemical modifications to DNA or to the proteins that are closely associated with DNA (the histones, which form the cores of chromatin packaging), and a prominent role for RNA is also emerging. Given this developing mechanistic understanding, the field is attracting investigators interested in diverse aspects of chromatin and chromosome biology.

In this Insight, we take a wide view of the epigenetics field, highlighting current topics of interest — from the influence of chromatin and chromosome organization on gene expression to the roles of epigenetic mechanisms in development and disease. And under this broad umbrella, the very definition of epigenetics is scrutinized. We hope that you enjoy these exciting reviews and thank the authors for their contributions.

We are pleased to acknowledge the financial support of Invitrogen and March of Dimes, which contributed towards the distribution of this Insight. As always, *Nature* carries sole responsibility for editorial content and peer review.

Alex Eccleston, Natalie DeWitt, Chris Gunter, Barbara Marte and Deepa Nath, Senior Editors

INTRODUCTION

REVIEWS

nature insight

NATURE|Vol 447|24 May 2007|doi:10.1038/nature05913

Perceptions of epigenetics

Adrian Bird[1]

Geneticists study the gene; however, for epigeneticists, there is no obvious 'epigene'. Nevertheless, during the past year, more than 2,500 articles, numerous scientific meetings and a new journal were devoted to the subject of epigenetics. It encompasses some of the most exciting contemporary biology and is portrayed by the popular press as a revolutionary new science — an antidote to the idea that we are hard-wired by our genes. So what is epigenetics?

There has always been a place in biology for words that have different meanings for different people. Epigenetics is an extreme case, because it has several meanings with independent roots. To Conrad Waddington, it was the study of epigenesis: that is, how genotypes give rise to phenotypes during development[1]. By contrast, Arthur Riggs and colleagues defined epigenetics as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence"[2]: in other words, inheritance, but not as we know it. These definitions differ markedly, although they are often conflated as though they refer to a single phenomenon. Waddington's term encompasses the activity of all developmental biologists who study how gene activity during development causes the phenotype to emerge, but it suffers from the disadvantage that developmental biologists themselves rarely, if ever, use this word to describe their field. In this sense, the usage is obsolete. The definition put forward by Riggs and colleagues tells us what epigenetics is not (inheritance of mutational changes), leaving open what kinds of mechanism are at work. In this article, I give examples of how epigenetic phenomena are studied and interpreted, and I propose a revised definition that embodies contemporary usage of the word.

The molecular basis of heritable epigenetics has been studied in a variety of organisms. The DNA methylation system and the Polycomb/Trithorax systems come closest to the ideal, because alterations in these systems are often inherited by subsequent generations of cells and sometimes organisms (Box 1). A classic case of what Robin Holliday named epimutation[3] is the peloric variant of toadflax (*Linaria*) flowers (Fig. 1), first described by Linnaeus. In this variant, heritable silencing of the gene *Lcyc*, which controls flower symmetry, is due not to a conventional mutation (that is, a mutation in the nucleotide sequence) but to the stable transmission of DNA methylation at this locus from generation to generation[4]. Although most variants arising in laboratory plants are due to conventional mutations rather than epimutations of this kind, examples of transgenerational epigenetics are now well documented in plants (see page 418) and fungi. In animals, however, the transmission of epigenetic traits between organismal generations has, so far, been detectable only by using highly sensitive genetic assays[5]. The mouse agouti locus (also known as nonagouti), which affects coat colour, is the best-studied example, being affected by the extent of DNA methylation at an upstream transposon. Genetically identical parents whose agouti genes are in different epigenetic states tend to produce offspring with different coat colours, although the effect is variable.

Despite the paucity of data from animal studies, this type of epigenetics has caught the general imagination because, in principle, it is stable but potentially affected by the environment. The possibility that acquired 'marks' can be passed from parents to children has a deliciously lamarckian flavour that has proved difficult to resist as a potential antidote to genetic determinism. A recent BBC television science programme hailed the advent of epigenetics as a profound shift in our understanding of inheritance (http://www.bbc.co.uk/sn/tvradio/programmes/horizon/ghostgenes.shtml). It summarized the implications of the emergent science as follows: "At the heart of this new field is a simple but contentious idea — that genes have a 'memory'. That the lives of your grandparents — the air they breathed, the food they ate, even the things they saw — can directly affect you, decades later, despite your never experiencing these things yourself." Is there any evidence for these heady claims, and how reliable is it? The answer to the first part of the question is yes.

Genes learning by experience?

Several studies have reported evidence that links the environment or ageing to long-lasting epigenetic effects on phenotype. One study

Box 1 | Epigenetic paradigms

There are two classic epigenetic systems: the Polycomb and Trithorax (Polycomb/Trithorax) systems, and DNA methylation. The Polycomb and Trithorax groups of proteins, which are named after mutants of the fruitfly *Drosophila melanogaster*, work to maintain repressed or active transcription states, respectively, of developmentally important genes. In the absence of these systems, the genes that specify the different segments of the fruitfly are initially expressed correctly, but this pattern cannot be maintained. It can be inferred from this that the Polycomb/Trithorax systems stably 'memorize' gene-expression patterns that have been set up by other cellular mechanisms. There is evidence that Polycomb-imposed silencing can even be transmitted between fruitfly generations at low frequency[18]. Biochemical studies have enabled the identification of components of the two key Polycomb-system protein complexes and have established a close link with modification of the lysine residue at position 27 of histone H3. The mechanism by which silencing is transmitted between cell generations remains obscure.

In the case of DNA methylation, biochemical information preceded genetic understanding of the system. The methylated sequence in vertebrates is CG, which is paired with the same sequence on the opposite DNA strand. This symmetry means that sites are transiently methylated on only one of the two DNA strands (that is, hemimethylated) after DNA replication. CG methylation patterns are copied between cell generations by the DNA methyltransferase DNMT1, which 'completes' hemimethylated but not unmethylated sites. In plants and fungi, the base 5-methylcytosine is also present in non-symmetrical DNA sequences, so the mechanism of copying is less obvious. DNA methylation is associated with stable gene silencing (for example, on the inactive X chromosome), either through interference with transcription-factor binding or through the recruitment of repressors that specifically bind sites containing methylated CG.

[1]Wellcome Trust Centre for Cell Biology, Edinburgh University, The King's Buildings, Edinburgh EH9 3JR, UK.

examined monozygotic (that is, identical) twins, whom, perhaps oddly, epigeneticists often use to exemplify their system at work. To many, twins epitomize the awesome power of genetics to determine human form and function regardless of environment. Indeed, 'concordance' of a particular characteristic in monozygotic and dizygotic twins is one of the most reliable ways of assessing its genetic basis. What has attracted the attention of epigeneticists, however, is that monozygotic twins do not always show the same disease susceptibility, raising the possibility that epigenetic differences that arise during ageing are at work[6]. Accordingly, it has been reported that young twins have similar amounts of DNA methylation, whereas older twins differ considerably in the amounts and patterns of this modification[7]. Might these non-genetic age-dependent differences in gene marking give rise to the divergent disease predispositions seen in some twins? At present, this is unclear, and a recent study emphasizes the need for further basic work on twins. The largest high-resolution analysis of human DNA methylation patterns so far found that 873 genes on 3 chromosomes showed no significant variation in DNA methylation between individuals in their mid-20s and those in their mid-60s[8]. The remarkable uniformity of DNA methylation among unrelated individuals of disparate ages does not square easily with the large divergence reported in twins of the same age.

Another high-profile study has raised the possibility that a mother's behaviour can affect the chemistry of DNA in her offspring. Quality of early maternal care has long been acknowledged to have long-term repercussions during the lifetime of an individual. A potential mechanism for this effect was deduced from a study reporting that maternal nurturing in rats alters DNA methylation at the gene encoding the glucocorticoid receptor[9]. The authors suggest that in the absence of appropriate nurturing, there is less methylation of this gene in the hippocampus, resulting in overexpression of the receptor in later life. The implication is that the glucocorticoid-mediated stress-response pathway is epigenetically fixed at the level of gene transcription. In addition, transgenerational effects of environmental insults have been reported in mammals: for example, the exposure of embryonic rats to the anti-androgenic compound vinclozolin led to a decrease in spermatogenesis not only in the treated animals but also in males of several subsequent generations[10]. Altered DNA methylation was again suggested as a potential mediator of this effect, although, during development, mammalian embryos pass through a profoundly hypomethylated state, which might be expected to jeopardize the heritability of such marks. Despite uncertainties about the mechanism(s) at work, these studies have raised the profile of epigenetics as a potential mechanistic explanation for the long-term impact of the environment on physiology and behaviour (see page 433). Time will tell whether that potential is realized.

Epigenetics and inheritance

Should heritability be mandatory in a contemporary view of epigenetics? The requirement that epigenetic characters should be transmissible through mitosis or meiosis has the virtue of clarity but can be a liability. To explain why, it is necessary to introduce a third, somewhat informal, 'definition' of epigenetics that has crept into widespread use. This incarnation of epigenetics encompasses the biology of chromatin, including the complex language of chromatin marks (see page 407), the transcriptional effects of RNA interference (see page 399) and, for good measure, the effects of the higher-order structure of chromosomes and the nucleus (see page 413). The attraction of this usage is that it brackets together some of the most exciting contemporary work in biology. Its drawback is that it does not sit easily with the prevailing textbook definitions. One reason for this is that many chromatin marks are short-lived. For example, phosphorylation of the variant histone H2AX (also known as H2AFX) after a double-strand break[11] would qualify as an epigenetic mark under the emerging definition, but it is too transient to qualify as a heritable epigenetic mark (Fig. 2). Histone modifications associated with transcription are also ambiguous with respect to heritability. On the one hand, DNA methylation affects histone acetylation and histone methylation, so these modifications can be viewed as heritably epigenetic, albeit indirectly[12]. On the other hand, these histone marks can also result from events that seem to involve neither DNA methylation nor Polycomb group proteins, and the marks are not necessarily transmissible between generations. Therefore, a single histone modification could, in principle, be rated as either epigenetic or not epigenetic according to the heritability credentials of its origin. Such a complicated classification system would have limited utility.

The issue of replicative accuracy is also relevant when considering heritability. DNA synthesis is spectacularly accurate, making only

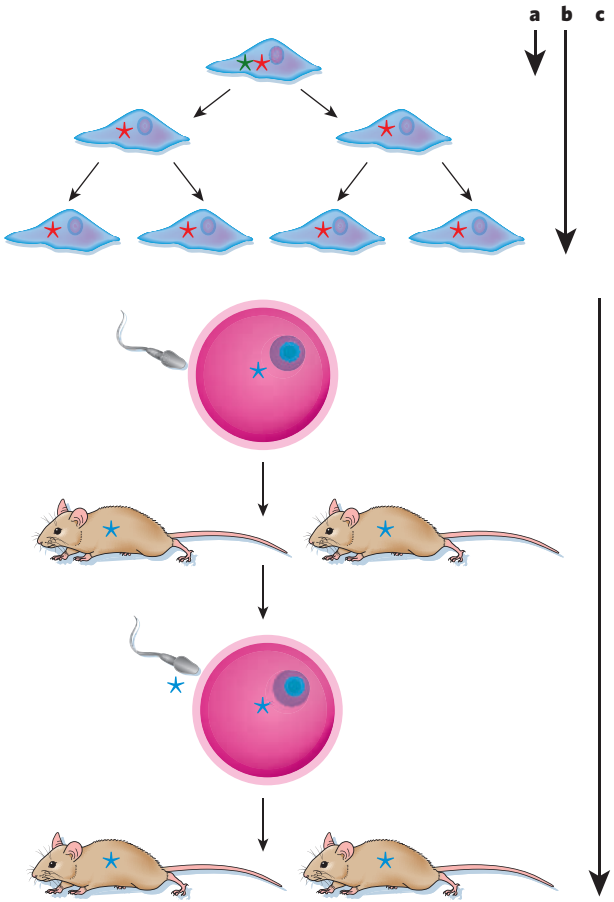


Figure 2 | Persistence of epigenetic marks. Alterations that last less than one cell cycle (green asterisk, **a**) do not qualify as epigenetic under the definition that strictly requires heritability, whereas non-mutational changes that are transmitted from one cell to its daughters (red asterisk, **b**) or between generations of an organism (blue asterisk, **c**) do qualify.

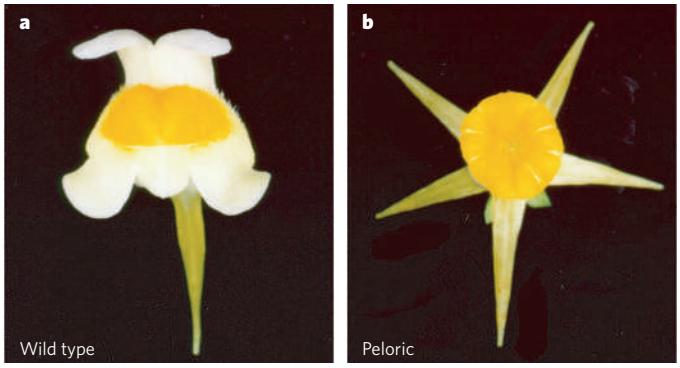


Figure 1 | Frontal view of a wild-type toadflax flower and a peloric epimutant. a, The wild-type flower is dorsoventrally asymmetrical. **b**, By contrast, the peloric flower is radially symmetrical with all petals resembling the ventral petal of the wild-type flower. (Image reprinted, with permission, from ref. 4.)

1 'unforced' error for every 10^7–10^8 bases copied[13]. But DNA methylation has an apparent accuracy of ~96%, which is ~1 error for every 25 methylated sites copied[14]. Because of this error rate, cloning from a single cell quickly results in a population of cells in which DNA methylation patterns are diverse[15]. Methylated domains are more stably maintained, even though the detailed location of methylated sites varies within them. But even the peloric variant of toadflax, which is an otherwise perfect example of heritable epigenetics in action, shows considerable instability as the plant grows. So how accurately transmitted should an epigenetic mark be? Variation due to faulty copying is compounded by current evidence that all histone modifications, as well as DNA methylation itself, can be abruptly removed during development, thereby preventing the persistence of these modifications in a heritable epigenetic sense (see page 425). The restrictiveness of the heritable view of epigenetics is perhaps best illustrated by considering the brain. A growing idea is that functional states of neurons, which can be stable for many years, involve epigenetic phenomena[16], but these states will not be transmitted to daughter cells because almost all neurons never divide.

Refining a definition

Given that there are several existing definitions of epigenetics, it might be felt that another is the last thing we need. Conversely, there might be a place for a view of epigenetics that keeps the sense of the prevailing usages but avoids the constraints imposed by stringently requiring heritability. The following could be a unifying definition of epigenetic events: the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states. This definition is inclusive of chromosomal marks, because transient modifications associated with both DNA repair or cell-cycle phases and stable changes maintained across multiple cell generations qualify. It focuses on chromosomes and genes, implicitly excluding potential three-dimensional architectural templating of membrane systems and prions, except when these impinge on chromosome function. Also included is the exciting possibility that epigenetic processes are buffers of genetic variation, pending an epigenetic (or mutational) change of state that leads an identical combination of genes to produce a different developmental outcome[17].

An implicit feature of this proposed definition is that it portrays epigenetic marks as responsive, not proactive. In other words, epigenetic systems of this kind would not, under normal circumstances, initiate a change of state at a particular locus but would register a change already imposed by other events. Such events could be, for example, the collision of DNA with ionizing radiation or a developmental switch in gene expression. It could be argued that the responsive nature of epigenetic processes is a unifying feature, because classic epigenetic systems such as the DNA methylation system and the Polycomb/Trithorax systems seem to respond to previous switches in gene activity in this way. Therefore, their sophisticated feature is the ability, in the 'darkness' of the nucleus, to sense and mark changes in the chromosomal status. For example, transcriptional activation through sequence-specific DNA-binding proteins brings in histone acetyltransferases, which then epigenetically adapt the promoter region for transcription (for histone acetyl groups, although ephemeral, would now be epigenetic). Similarly, elongating polymerases carry enzymes that restrain the spurious transcriptional initiation that might arise within the temporarily disrupted chromatin of an active gene. Without such epigenetic mechanisms, hard-won changes in genetic programming could be dissipated and lost; transient disruptions of chromosomal organization might go uncompensated; and DNA damage might escape repair. ■

1. Waddington, C. H. *The Strategy of the Genes* (Allen & Unwin, London, 1957).
2. Russo, V. E. A., Martienssen, R. A. & Riggs, A. D. (eds) *Epigenetic Mechanisms of Gene Regulation* (Cold Spring Harbor Laboratory Press, Woodbury, 1996).
3. Jeggo, P. A. & Holliday, R. Azacytidine-induced reactivation of a DNA repair gene in Chinese hamster ovary cells. *Mol. Cell. Biol.* **6,** 2944–2949 (1986).
4. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401,** 157–161 (1999).
5. Chong, S. & Whitelaw, E. Epigenetic germline inheritance. *Curr. Opin. Genet. Dev.* **14,** 692–696 (2004).
6. Wong, A. H., Gottesman, I. I. & Petronis, A. Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum. Mol. Genet.* **14,** R11–R18 (2005).
7. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* **102,** 10604–10609 (2005).
8. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genet.* **38,** 1378–1385 (2006).
9. Weaver, I. C. *et al.* Epigenetic programming by maternal behavior. *Nature Neurosci.* **7,** 847–854 (2004).
10. Anway, M. D., Cupp, A. S., Uzumcu, M. & Skinner, M. K. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308,** 1466–1469 (2005).
11. Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. *J. Cell Biol.* **146,** 905–916 (1999).
12. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31,** 89–97 (2006).
13. Kunkel, T. A. DNA replication fidelity. *J. Biol. Chem.* **279,** 16895–16898 (2004).
14. Laird, C. D. *et al.* Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc. Natl Acad. Sci. USA* **101,** 204–209 (2004).
15. Silva, A. J., Ward, K. & White, R. Mosaic methylation in clonal tissue. *Dev. Biol.* **156,** 391–398 (1993).
16. Hong, E. J., West, A. E. & Greenberg, M. E. Transcriptional control of cognitive development. *Curr. Opin. Neurobiol.* **15,** 21–28 (2005).
17. Sollars, V. *et al.* Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nature Genet.* **33,** 70–74 (2003).
18. Cavalli, G. & Paro, R. The *Drosophila Fab-7* chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93,** 505–518 (1998).

Acknowledgements I thank the Wellcome Trust for research support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (apbird@staffmail.ed.ac.uk).

Transcription and RNA interference in the formation of heterochromatin

Shiv I. S. Grewal[1] & Sarah C. R. Elgin[2]

Transcription in heterochromatin seems to be an oxymoron — surely the 'silenced' form of chromatin should not be transcribed. But there have been frequent reports of low-level transcription in heterochromatic regions, and several hundred genes are found in these regions in *Drosophila*. Most strikingly, recent investigations implicate RNA interference mechanisms in targeting and maintaining heterochromatin, and these mechanisms are inherently dependent on transcription. Silencing of chromatin might involve *trans*-acting sources of the crucial small RNAs that carry out RNA interference, but in some cases, transcription of the region to be silenced seems to be required — an apparent contradiction.

Chromatin fibres, which make up chromosomes, are composed of nucleosome arrays, with each nucleosome consisting of an octamer of core histones associated with double-stranded DNA. Great variety in chromatin biochemistry is achieved by a complex system of accessory proteins, which modify, bind and reorganize histone complexes to generate different functional regions in eukaryotic chromosomes. Chromatin can be considered to have two main types of domain: euchromatin, which is gene-rich; and heterochromatin, which is gene-poor. These domains have different patterns of histone modification, are associated with different modes of nucleosome packaging[1] and therefore, presumably, have differences in higher-order packaging[2,3] and nuclear organization (see page 413).

Heterochromatin was initially defined as the portion of the genome that retains deep staining with DNA-specific dyes as the dividing cell returns to interphase from metaphase. Subsequent investigation showed that heterochromatin has a constellation of properties (Box 1). A link between heterochromatin formation and gene silencing has been inferred from the loss of most gene activity on the inactive X chromosome, which is visibly condensed in female mammals, and from the loss of gene expression, correlated with condensed packaging, in position-effect variegation (PEV) in *Drosophila* and other organisms. PEV occurs when a gene that is normally euchromatic is juxtaposed with heterochromatin, through rearrangement or transposition; the resultant variegating phenotype indicates that the gene has been silenced in a proportion of the cells in which it is normally active[1]. Reporter genes that show PEV (packaged in heterochromatin) have a more uniform nucleosome array and, perhaps as a consequence, suffer a loss of 5′ nuclease-hypersensitive sites (that is, regions that are presumed to be nucleosome free and are generally associated with regulatory sequences present in active or readily induced genes)[2,4]. Loss of nuclease-hypersensitive sites depends on Heterochromatin protein 1 (HP1; also known as Suppressor of variegation 205, SU(VAR)205)[5]. Studies in fission yeast, *Schizosaccharomyces pombe*, have shown that HP1-family proteins mediate recruitment and/or spreading of chromatin-modifying factors, such as the multi-enzyme complex SHREC (SNF2- and histone deacetylase (HDAC)-containing repressor complex). Such complexes presumably facilitate the nucleosome modification and positioning needed to organize the higher-order chromatin structures that are essential for diverse heterochromatin functions, including silencing of transcription, suppression of recombination, long-range chromatin interactions and maintenance of genomic integrity[1,3,6].

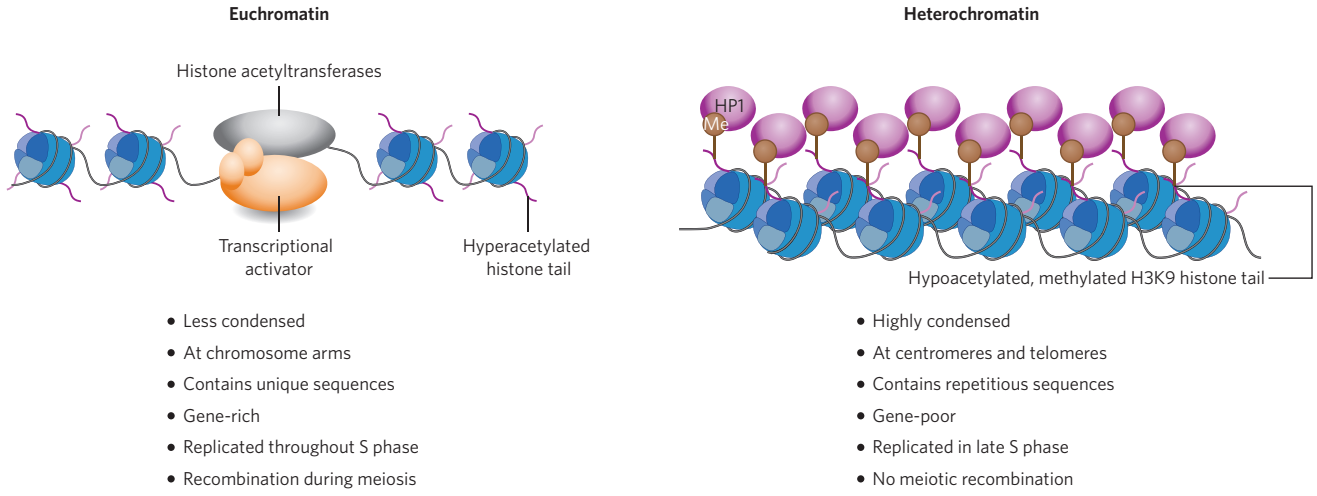
An important characteristic of heterochromatin is the ability of this form of packaging to spread, as evidenced by the occurrence of PEV in *Drosophila* and as shown in *S. pombe*[7] (discussed later). After heterochromatin has been established, it can be stably maintained through mitosis, as shown by the patchy coat of the tortoiseshell (calico) cat, in which coat colour depends on which X chromosome is inactivated. The general properties of euchromatin are antagonistic to those of heterochromatin (Box 1), although it is anticipated that, in euchromatin, there is much more variation in the modification state of the histones and in the arrangement of the nucleosome array, both of which depend on the transcriptional state of a given gene (see page 407). Indeed, greater expression of a *Drosophila* gene (embedded in heterochromatin) that confers variegation has been reported in response to introducing increased amounts of a transcription factor (Gal4), suggesting that there is constant competition in establishing these alternative states[8]. Furthermore, despite the clear distinctions between heterochromatin and euchromatin, low-level transcription has often been found to occur in heterochromatic regions, and these regions contain several hundred genes in *Drosophila*[9]. Resolving these apparent contradictions will provide new insight into how genomes function. In this review article, we focus on recent findings about how heterochromatin formation is targeted and maintained in specific regions of the genome, examining the potential role of transcription associated with the RNA interference (RNAi) system. We draw mainly on results from studies of fungi and animals; interesting results from plants are reported on page 418.

Heterochromatin assembly in *Drosophila*

A key tool for investigating heterochromatin has been the ability to screen for suppressors of PEV (*Su(var)*): that is, mutations elsewhere in the genome that result in loss of silencing at a variegating locus. About 15 such loci have been characterized in *Drosophila melanogaster*, and many more candidates have been identified[10]. The *Su(var)* genes typically encode either proteins that participate directly in the structure of heterochromatin or enzymes that control changes in the modification of histones. A transition between euchromatin and heterochromatin (as might occur in PEV) can roughly be viewed as a series of reactions in which the histone modifications and the proteins associated with the

[1]Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [2]Washington University, Department of Biology CB1137, St. Louis, Missouri 63130, USA.

Box 1 | Properties of euchromatic and heterochromatic regions



- Less condensed
- At chromosome arms
- Contains unique sequences
- Gene-rich
- Replicated throughout S phase
- Recombination during meiosis

- Highly condensed
- At centromeres and telomeres
- Contains repetitious sequences
- Gene-poor
- Replicated in late S phase
- No meiotic recombination

Trying to define heterochromatin is like trying to define life itself: a cluster of important properties can be specified, but there are exceptions in every instance. For example, centromeres are usually associated with blocks of flanking heterochromatin. However, the inner centromere of *Drosophila* chromosomes is associated with blocks of nucleosomes that contain CENP-A (also known as CID), a variant of histone H3, interspersed with blocks of nucleosomes that contain H3 with a different modification pattern[78]. The elements at the telomeres of *Drosophila* chromosomes are non-LTR (non-long terminal repeat) retrotransposons, which are transcribed[79]. By contrast, the proximal telomere-associated sequences show more of the properties of heterochromatin. The characteristics listed in the figure are most consistently observed in pericentromeric heterochromatin: that is, in the regions that flank the centromeres of many eukaryotic chromosomes. These regions are rich in remnants of transposable elements. It should be noted that little is known about either the stoichiometry of HP1 or the folding of the chromatin fibre in heterochromatin; the figure is meant to convey only the association of HP1 and the condensation of the chromatin fibre.

active state are removed, and then the histone modifications and proteins associated with the inactive state are added (Fig. 1). A sequential set of reactions is required: for example, the lysine residue at position 9 of histone H3 (H3K9) cannot be methylated until it is deacetylated; the binding of SUV4-20 to heterochromatic loci occurs through interaction with HP1 and requires the activity of SU(VAR)3-9 (ref. 11). This sequential requirement undoubtedly contributes to the relative stability of the alternative packaging states. Although the heterochromatic state can be inherited through mitotic and even meiotic cell divisions, a given site can switch from a repressed to an active chromatin state and vice versa at a low frequency. PEV cannot be scored this way in single-celled organisms such as yeast, but this switching can be observed in the phenotype of sectors of a growing colony (Fig. 2).

In *Drosophila*, a small group of proteins are considered likely structural components of pericentromeric heterochromatin because of the observed dose response: whereas having one copy of the encoding gene results in loss of silencing, having three copies results in increased silencing, presumably due to mass action[12]. This set of proteins includes the following: HP1, the first chromodomain protein to be identified[13]; HP2 (also known as SU(VAR)2-HP2), a large protein with no conserved structural motifs[14]; SU(VAR)3-7, a zinc-finger protein[15]; and SU(VAR)3-9, an H3K9-specific histone methyltransferase[16]. However, a well-defined complex of these heterochromatin-associated proteins has not been isolated (despite HP1 interacting with all of these proteins and other proteins[17]), suggesting that an organized protein assembly is present only on the chromatin fibre.

HP1 is a small protein (206 amino acids in *D. melanogaster*) with two conserved domains, an amino-terminal chromodomain and a carboxy-terminal chromoshadow domain, separated by a hinge region (Fig. 3). The chromodomain, found in many chromosomal proteins, folds to create a binding site for the N-terminal tails of histones. HP1 dimerizes through the chromoshadow domain, forming a peptide-binding surface. HP1 interacts stably with SU(VAR)3-9 through the chromoshadow and hinge domains and with di- or trimethylated H3K9 (H3K9me2 or H3K9me3) through the chromodomain[1]. By interacting with both histone-modifying enzyme and modified histone, HP1 provides a foundation for a self-assembly and spreading mechanism, which has been anticipated from studies of PEV (Fig. 3) (see ref. 18 for a review of possible spreading mechanisms). This core assembly seems to be conserved across animals and fungi[19,20] (Table 1). It should be noted that, in many organisms, there are several homologues of HP1 and multiple H3K9 methyltransferases, suggesting the possibility of alternative protein assemblies[19]. However, in *Drosophila*, only HP1A (referred to as HP1 in this review) seems to be associated with known heterochromatic regions, and the ability of other homologues to mimic HP1 in establishing heterochromatin packaging remains to be determined.

The role of RNAi in *S. pombe*

Genetic and biochemical studies using *S. pombe* as a model system have provided great insight into the mechanisms of heterochromatin assembly. Many of the factors involved in heterochromatin formation in *Drosophila* and mammals are conserved in *S. pombe*[19,20]. In particular, the protein Clr4 (cryptic loci regulator 4) — which is the *S. pombe* homologue of *Drosophila* SU(VAR)3-9 and is present in an E3–ubiquitin ligase complex that contains cullin 4 (also known as Pcu4) — has been shown to methylate H3K9 specifically[21–26]. Methylated H3K9 functions as a binding site for recruitment of chromodomain-containing proteins — including chromodomain protein 1 (Chp1), Chp2 and Swi6 (the last of which is a homologue of *Drosophila* HP1) — to heterochromatic loci[22,27–29]. Heterochromatin-associated factors, including methylated H3K9 and Swi6, were found to map to extended chromosomal regions that are coated with heterochromatin complexes at centromeres, telomeres and the mating-type locus[30]. Interestingly, all three of these heterochromatic regions have a common feature — each contains *dg* and *dh* repeat elements, which are preferential targets of heterochromatin formation[7,31–33]. Recent investigations into mechanisms by which these repeats might trigger heterochromatin formation led to the surprising discovery that the RNAi system is involved in the nucleation and assembly of heterochromatin[7,33].

RNAi was first described as a post-transcriptional silencing mechanism in which double-stranded RNA triggers the destruction of cognate RNAs[34]. Subsequent studies have implicated RNAi-associated

mechanisms in diverse cellular functions. In *S. pombe*, mutations in genes encoding factors that are involved in RNAi — such as dicer (Dcr1; an enzyme that cleaves double-stranded RNA), argonaute (Ago1; a PAZ- and PIWI-domain-containing protein that can bind small RNAs) and RNA-directed RNA polymerase 1 (Rdp1) — result in defects in heterochromatin assembly, as shown by loss of silencing at reporter loci[7,33]. An RNAi-induced transcriptional silencing complex (RITS), which contains both a chromatin-associated protein and an RNAi-associated protein, has been identified[35]. RITS contains Chp1 (a chromodomain protein), Ago1 and a protein of unknown function, Tas3 (RITS subunit 3). In addition, RITS also contains small interfering RNAs (siRNAs) derived from the *dg* and *dh* repeats present at the different heterochromatic loci[30,35]. Genome-mapping analyses have shown that Rdp1 and components of RITS are distributed throughout heterochromatic regions in a pattern that is almost identical to the distribution of Swi6 and of H3K9 methylation[30]. Stable binding of RITS to chromatin depends, at least in part, on the binding of the chromodomain of Chp1 to methylated H3K9 (ref. 36). Deletion of *clr4*, or a mutation in the chromodomain-encoding region of *chp1*, results in delocalization of RITS from heterochromatic loci. Interestingly, there are concurrent defects in the processing of *dg* and *dh* repeat transcripts into siRNAs[30,36], suggesting that siRNAs are produced in a heterochromatic environment.

RITS also recruits an RNA-directed RNA polymerase complex (RDRC) that contains Rdp1; this polymerase activity is essential for siRNA production and heterochromatin assembly[37,38]. The generation of siRNAs also requires an RNaseH-like RNA-cleavage activity (referred to as slicer activity) known to be associated with argonaute-family proteins, such as Ago1, found in RITS. Mutations in conserved Ago1 residues that abolish this activity severely affect the processing of *dg* and *dh* repeat transcripts and result in defects in heterochromatin assembly[39,40]. The slicer function of Ago1 has been suggested to be important for the spreading of heterochromatin[39]. It is also possible that siRNAs generated by Ago1-mediated processing of transcripts have a direct structural role in the assembly of higher-order structures that, in addition to mediating silencing, facilitates the local spreading of heterochromatin. These mechanisms, however, cannot by themselves account for the spreading of heterochromatin across large regions, because this requires the HP1-family protein Swi6, which functions as a platform for recruiting the chromatin-modifying effectors (that is, proteins or complexes) involved in heterochromatin assembly[1,7].

These findings suggest that RNAi-mediated heterochromatin assembly in *S. pombe* might occur through a self-reinforcing loop[36,37]. In this model, siRNAs (possibly generated elsewhere) and/or DNA-binding proteins mediate the initial targeting of heterochromatin-associated factors, resulting in the establishment of H3K9 methylation. The presence of methylated H3K9 and associated silencing factors, in turn, allows stable binding of RITS across heterochromatic regions (Fig. 4). RITS presumably functions as a core for the binding of other RNAi-associated factors, such as RDRC, that are essential for the processing of any *dg* and *dh* repeat transcripts. The siRNA-guided cleavage of nascent repeat transcripts by Ago1 (a component of RITS) is thought to be an important step in producing additional siRNAs. It is possible that cleaved transcripts are preferential targets for Rdp1. Rdp1 generates double-stranded RNAs, which are necessary for the generation of siRNAs by Dcr1. Those siRNAs produced *in cis* can feed back to target more heterochromatin complexes but might also have other functions (discussed in the next section).

The exact mechanism by which siRNAs target histone modifications is unclear. The binding of RITS to heterochromatic regions requires *dg* and *dh* siRNAs to be part of the complex. It has been suggested that RITS, tethered to nascent transcripts by siRNAs, might mediate the recruitment of histone methyltransferases such as Clr4 (ref. 35) or that siRNAs directly facilitate the recruitment of chromatin-modifying effectors, such as the Clr4-containing complex, to heterochromatic repeats[7,23]. It is certainly possible that siRNAs target heterochromatin by base-pairing with nascent transcripts[41]; subunits of RITS and RDRC can be crosslinked to transcripts of non-coding centromeric repeats[38]. However, it is unknown whether this binding simply reflects the roles of these factors in processing repeat transcripts or whether it indicates an additional function in recruiting heterochromatin proteins. Recently, artificial tethering of RITS to nascent transcripts has been shown to induce

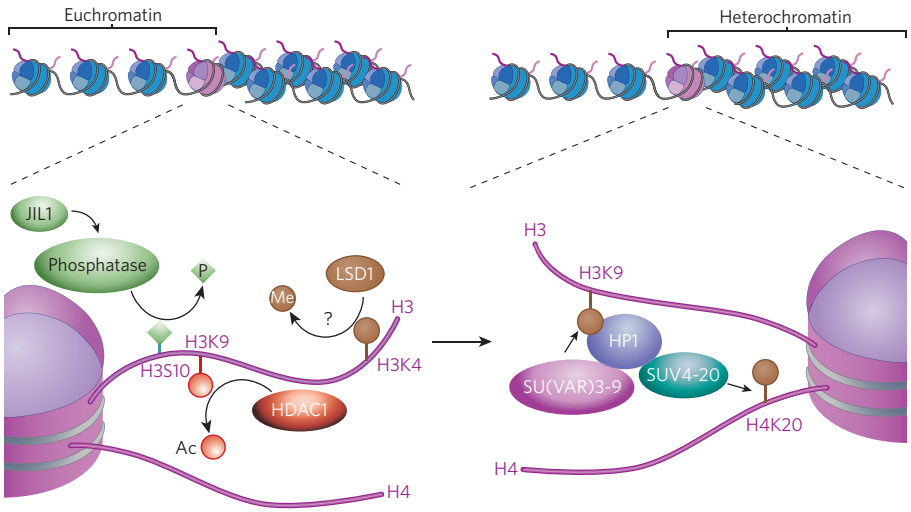


Figure 1 | Changes in histone modification implicated in the switch from a euchromatic to a heterochromatic state in *Drosophila*. Active genes are frequently marked by H3K4me3 (see page 407); this modification is presumably removed by LSD1 (which has not yet been characterized in *Drosophila*). H3K9 is normally acetylated in euchromatin, and this modification must be removed by a histone deacetylase, typically HDAC1. Phosphorylation of H3S10 can interfere with the methylation of H3K9; its dephosphorylation might involve a phosphatase targeted through the carboxy terminus of the protein kinase JIL1 (ref. 10). These transitions set the stage for acquisition of the modifications that are associated with silencing: these include the methylation of H3K9 by SU(VAR)3-9 or another histone methyltransferase, the binding of HP1, and the subsequent methylation of H4K20 by SUV4-20 (an enzyme that is recruited by HP1). Other silencing marks such as methylation of H3K27 by E(Z) (enhancer of zeste; not shown) seem to be relevant in some regions, although this mark is more prominently used by the Polycomb system. Supporting data come from genetic identification of modifiers of PEV, as well as biochemical characterization of the activities of such modifiers and tests of protein–protein interactions[10]. (Figure adapted, with permission, from ref. 10.)

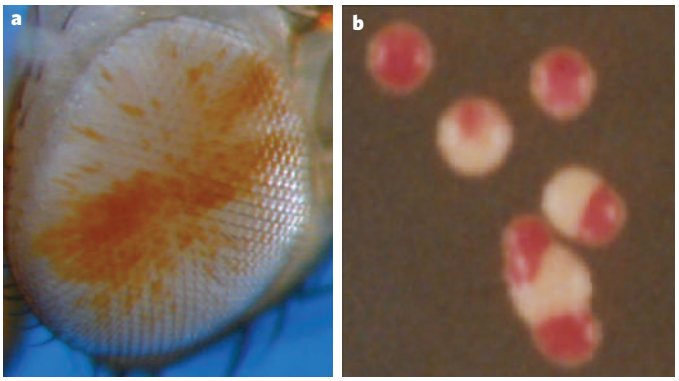


Figure 2 | Variegating phenotypes. Although alternative chromatin packaging states (that is, euchromatin and heterochromatin) can be inherited, they switch at a low frequency. This results in a variegating phenotype in a clonal population of cells. **a**, The image shows a *Drosophila* eye. The *white* gene, expression of which results in a red eye, is active in some eye facets but silenced in others. (Image courtesy of E. Gracheva, Washington University in St. Louis, Missouri.) **b**, The image shows colonies of the fission yeast, *S. pombe*, each of which has differently coloured sectors as a result of variegated expression of the *ade6* gene inserted in a heterochromatic region. (Image courtesy of K. Noma, National Cancer Institute, National Institutes of Health, Bethesda, Maryland.)

local heterochromatin assembly[42]. However, this process requires Dcr1, presumably for the production of siRNAs. Therefore, in addition to the targeting of RITS, other siRNA-dependent steps are required for stable RNAi-mediated heterochromatin nucleation. The emerging view is that, through associations with components of heterochromatin, the RNAi machinery — tethered to specific loci — helps to process transcripts generated from these loci into siRNAs, thereby effectively causing post-transcriptional silencing *in cis*[30,36]. The siRNAs produced in this process also facilitate further targeting of heterochromatin modifications, such as H3K9 methylation. H3K9 methylation enables HP1-family proteins such as Swi6 to localize across heterochromatic regions; these proteins, in turn, facilitate the localization of effectors (such as SHREC) with diverse cellular activities. The HDAC and ATPase (SNF2-like) activities of SHREC are crucial for the proper positioning of nucleosomes to achieve transcriptional silencing[3]. But how can transcription that generates siRNAs occur in a silenced region?

Transcription of heterochromatic repeats

From the results described in the previous section, it can be argued that heterochromatic repeats need to be transcribed to generate the siRNAs that target heterochromatin formation — a circular process. In support of this idea, recent studies have shown that heterochromatic repeats present in the *S. pombe* genome are transcribed by RNA polymerase II (Pol II)[30,43] and that mutations in Pol II impair RNAi-mediated heterochromatin assembly[43,44]. However, an apparent paradox arises in that heterochromatin, in general, is thought to be relatively inaccessible to factors involved in various aspects of DNA metabolism, including the transcriptional machinery[1]. How does Pol II gain access to sequences that are packaged as heterochromatin? Because heterochromatic silencing is thought to be plastic and can be overcome by an increased concentration of transcription factors[8], it can be argued that the promoters driving the transcription of repeats, unlike the promoters of euchromatic genes, have evolved to be somewhat impervious to heterochromatic repression. Indeed, one strand of centromeric repeats in *S. pombe* is always transcribed at a low level[33] but is silenced post-transcriptionally by RNAi-mediated processing of transcripts[33,36].

The transcription of repeats might be facilitated by a specialized mechanism(s) that modulates heterochromatin to provide access for factors involved in different chromosomal processes. In *S. pombe*, Swi6 (*Drosophila* HP1) is thought to function as an 'oscillator' of heterochromatin transcription by directing recruitment of both silencing and antisilencing factors[20]. In addition to factors (such as SHREC) that repress Pol-II-mediated transcription[3], Swi6 also recruits the JmjC-domain-containing protein Epe1 (ref. 40), which was identified in a screen for factors that negatively regulate heterochromatic silencing[45]. Epe1 facilitates Pol-II-mediated transcription of repeats specifically in the context of heterochromatin. It does not seem to have an obligatory role in transcription *per se*, because it is dispensable when heterochromatin is disrupted[40]. The mechanism by which Epe1 counteracts heterochromatic silencing is unknown. Because several JmjC-domain-containing proteins have been shown to catalyse histone demethylation[46], it is possible that Epe1 affects heterochromatin stability through the removal of repressive methylation of lysine residues. However, no such activity has been detected for Epe1 (ref. 47). Epe1 could modulate chromatin by an as yet undefined mechanism. Additional factors targeted to heterochromatic loci by Swi6 or by other mechanisms are also probably important for the transcription of heterochromatin.

In addition to heterochromatin assembly, the transcription of repeats embedded in heterochromatic regions probably has other biological implications. It has been suggested that the transcription of heterochromatic repeats is necessary for continuous production of siRNAs that prime the RNA-induced silencing complex (RISC)-like complexes required to neutralize future invasions by similar sequences[20]. The role of RNAi in destroying viral or transposable element transcripts is conserved in other species, including in *Tetrahymena thermophila* and *Drosophila*[48,49]. Heterochromatin-bound RNAi-associated factors might be components of a memory mechanism that selectively generates a reservoir of siRNAs directed against parasitic DNA elements[20]. It should be noted that, in *S. pombe*, RNAi machinery that is targeted to specific elements can spread to surrounding sequences, including nearby genes, by a process that depends on H3K9 methylation and Swi6 (ref. 36). This might also enable the RNAi machinery to exert heritable control over the expression of sequences located adjacent to repeats.

Silencing of repetitious sequences in metazoans

Although heterochromatin composed of repetitious DNA has become an essential part of the eukaryotic chromosome, maintaining the repetitious sequences in a stable, silent form (repressing both transposition and recombination) is clearly a challenge and a necessity. After it has been initiated, the packaging of heterochromatin occurs in a self-reinforcing manner, through multiple feedback loops[50]. The RNAi machinery seems to be able to detect and respond to repetitious DNA in a variety of ways. But, in metazoans, to what extent might RNAi components be used to target silent regions initially or to maintain these regions? And to what extent might silencing of repetitious DNA depend on transcription *in cis*? The system described in the previous section is unlikely to be universally applicable, because many metazoans, including *Drosophila* and mammals, seem to lack a canonical RNA-dependent RNA polymerase[19].

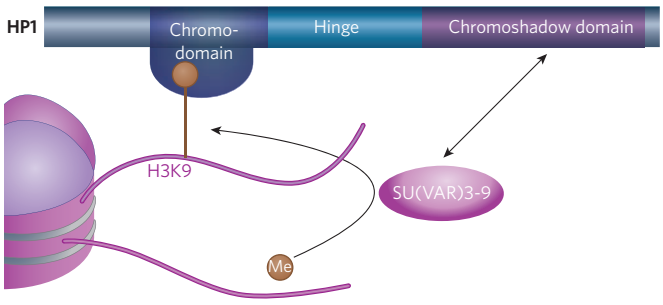


Figure 3 | HP1 and its interactions. HP1 interacts with H3K9me2 and H3K9me3 through its chromodomain, and with SU(VAR)3-9 through its chromoshadow domain. By interacting with both the modified histone and the enzyme responsible for the histone modification, HP1 provides a foundation for heterochromatin spreading and epigenetic inheritance. (Figure adapted, with permission, from ref. 10.)

Table 1 | Factors implicated in heterochromatin formation

Component	*S. pombe*	*Neurospora*	*Drosophila*	Mouse	*Arabidopsis*
Repetitious DNA	Yes	Yes	Yes	Yes	Yes
DNA methylation	No	Yes	No*	Yes	Yes
H3K9 methylation	Yes	Yes	Yes	Yes	Yes
HP1	Yes	Yes	Yes	Yes	No*
Small RNAs	Yes	No*	Yes	Yes	Yes
Pol II	Yes	ND	ND	ND	ND
RDR	Yes	No*	No	No	Yes

Yes indicates that the factor has been implicated to have a role in heterochromatin formation in the given organism. No indicates that the factor is not present in the organism. No* indicates that the organism has the factor but that it seems not to have a role in heterochromatin formation. ND means that the organism has the factor but whether it has a role in heterochromatin formation is unknown. *Arabidopsis*, *Arabidopsis thaliana*; *Neurospora*, *Neurospora crassa*; RDR, RNA-dependent RNA polymerase. (Table adapted from ref. 19.)

Post-transcriptional gene silencing mediated by RNAi, either the degradation of mRNA or a block in its translation, is known to occur in all metazoans that have been examined so far. The first suggestion of RNAi-based transcriptional gene silencing in *Drosophila* came from work showing a loss of expression when multiple copies of a transgene are present[51]. Subsequent analysis showed suppression of PEV (that is, a loss of silencing; as monitored through tandem arrays of *mini–white* and through *white* transgenes in heterochromatin) as a result of mutations in factors involved in the RNAi pathway[51–53]. The loss of silencing is associated with decreased levels of H3K9me2 (ref. 52). Similarly to the RNAi system in other organisms (notably plants; see page 418), the system in *D. melanogaster* might have originated as an antiviral defence mechanism[54,55]. About one-third of the genome is considered to be heterochromatic, and much of that DNA consists of remnants of transposable elements, both DNA transposons and retroviruses. The *Drosophila* genome encodes five PAZ- and PIWI-domain-containing proteins — PIWI, aubergine (AUB), AGO1, AGO2 and AGO3 — which are thought to bind small RNAs. PIWI is required for the self-renewal of germline stem cells, apparently having a key role in silencing retrotransposons and blocking their mobilization in the germ line[56]. Both PIWI and AUB are found associated with siRNAs of 24–29 nucleotides that are derived from repetitious sequences in the germ line[49,57]. *In vitro*, PIWI has RNA-cleavage activity[57], and it has been suggested that germline siRNAs might be generated by a unique processing mechanism that depends on cleavage of long single-stranded transcripts rather than double-stranded RNA[49]. How this silencing activity might influence heterochromatin formation in somatic cells (if at all) is unclear at present.

The effects of mutations in *AGO2* are clearly seen in early *Drosophila* embryos as defects in chromosome condensation, nuclear kinesis and spindle assembly, all potentially correlated with defects in the formation of centric heterochromatin[58]. Similar defects are observed when heterochromatin fails to form in *S. pombe* and other species[59–61]. In *Drosophila* with mutations in the genes that encode SU(VAR)3-9, HP1 or DCR-2, cells have disorganized nucleoli, as well as disorganized centric heterochromatin. In these circumstances, there is a substantial increase in extrachromosomal repetitious DNA in mutant tissues[62]. Similarly, mutations in the genes encoding the RNAi machinery in *S. pombe* also result in defects in maintaining chromosome integrity, including high rates of recombination at genes that encode ribosomal RNA[30]. Therefore, although repetitious DNA now contributes to essential chromosome structures, it is crucial to maintain this DNA specifically in a heterochromatic form, and genetic analysis indicates that the RNAi system has a role in this process. In the absence of any recognizable RNA-directed RNA polymerase activity, this might be accomplished by targeting heterochromatin formation to specific sites, either through DNA–protein interactions or through an RNAi-based recognition system, followed by spreading of the heterochromatin modifications and structure. Similarly to *S. pombe*, the spreading of heterochromatin in *D. melanogaster* (as monitored by PEV) depends on HP1 and SU(VAR)3-9.

Targeting heterochromatin formation

Although much of our discussion focuses on RNAi-based mechanisms, it is important to note that heterochromatin proteins can be recruited to specific sites (known as silencers) by DNA-binding factors. For example, in addition to the RNAi-mediated targeting of heterochromatin to a *dg*- and *dh*-like repeat element located in the silent mating-type region of *S. pombe*, the DNA-binding proteins Atf1 (activating transcription factor 1) and Pcr1, which belong to the ATF/CREB (cyclic-AMP-responsive-element-binding protein) family, have been shown to cooperate with components of SHREC to nucleate heterochromatin assembly independently in this region[6,63,64]. Similarly, redundant mechanisms of heterochromatin nucleation also operate at telomeres in *S. pombe*, where the TRF (TTAGGG repeat factor)-family DNA-binding protein Taz1, in conjunction with Ccq1 (coiled-coil quantitatively enriched protein 1), functions in parallel to the RNAi machinery to nucleate heterochromatin[3,32]. Regardless of the nucleation mechanism, heterochromatin targeted to specific sites can spread, and it provides a sequence-independent platform for cellular effectors with appropriate activities (such as SHREC, the RNAi machinery and cohesin) to be recruited across large regions[20].

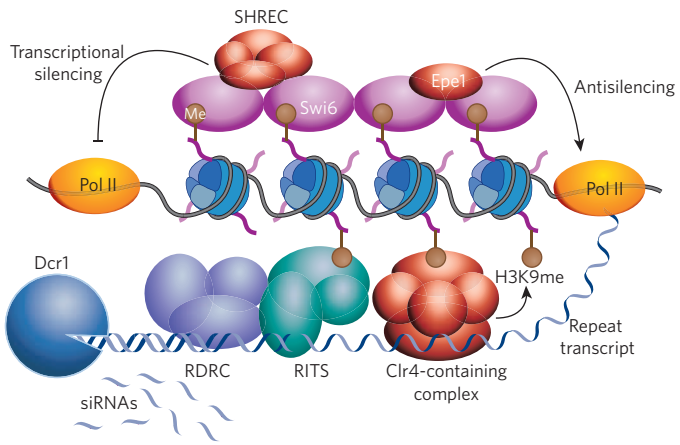


Figure 4 | Model showing RNAi-mediated heterochromatin assembly and silencing in *S. pombe*. Centromeric repeat (*dg* and *dh*) transcripts produced by Pol II are processed by the RNAi machinery, including the complexes RITS and RDRC (which interact with each other and localize across heterochromatic regions). The slicer activity of Ago1 (a component of RITS) and the RNA-directed RNA polymerase activity of Rdp1 (a component of RDRC) are required for processing the repeat transcripts into siRNAs. The siRNA-guided cleavage of nascent transcripts by Ago1 might make these transcripts preferential substrates for Rdp1 to generate double-stranded RNA, which in turn is processed into siRNAs by Dcr1. The targeting of histone-modifying effectors, including the Clr4-containing complex, is thought to be mediated by siRNAs. This process most probably involves the base-pairing of siRNAs with nascent transcripts, but the precise mechanism remains undefined. siRNAs produced by heterochromatin-bound RNAi 'factories' might also prime the assembly of RISC-like complexes capable of mounting a classic RNAi response. Methylation of H3K9 by Clr4 is necessary for the stable association of RITS with heterochromatic loci, apparently through binding to the chromodomain of Chp1. This methylation event also recruits Swi6, which, together with other factors, mediates the spreading of various effectors, such as SHREC. SHREC might facilitate the proper positioning of nucleosomes to organize the higher-order chromatin structure that is essential for the diverse functions of heterochromatin, including transcriptional gene silencing. Swi6 also recruits an antisilencing protein, Epe1, that modulates heterochromatin to facilitate the transcription of repeat elements, in addition to other functions. A dynamic balance between silencing and antisilencing activities determines the expression state of a locus within a heterochromatic domain.

In several cases documented in mammalian cells, HP1 can be targeted to specific promoters by interaction with DNA-binding complexes and seems to contribute to silencing at these loci (see ref. 65 for an example). However, in these cases, a different histone methyltransferase seems to be responsible for the accompanying H3K9 methylation, and spreading is generally not observed. These findings suggest that the interactions of HP1 with both the modified histone (that is, H3K9me3) and the modifying enzyme (usually SU(VAR)3-9) are crucial for heterochromatin spreading (Fig. 3).

Repetitious DNA is a hallmark of heterochromatin. In the case of satellite DNA (simple sequence tandem repeats), it can be suggested that a specific DNA-binding protein recognizes the satellite DNA sequence, thereby triggering heterochromatin assembly. In *Drosophila*, the protein D1, which when mutated results in a loss of silencing, preferentially binds satellite III DNA, which is (A+T)-rich[66,67]. Similarly, the heterochromatin-associated protein DDP1 (dodeca-satellite-binding protein 1; also known as DP1) binds to a conserved dodeca-satellite DNA sequence; however, this protein, which has 15 tandemly organized KH domains, also binds strongly to single-stranded nucleic acids with this sequence, including RNA[68]. Recent work has shown that DDP1, which also causes a loss of silencing when mutated, has a crucial role in the deposition of HP1 and methylated H3K9 at centromeric heterochromatin[68]. Given the ability of DDP1 (and its mammalian homologues, the vigilins) to bind RNA, it is possible that RNA mediates this interaction. Except for the blocks of satellite DNA, the repetitious sequences present in the *Drosophila* genome (which are mainly remnants of transposable elements and DNA transposons) are diverse. Consequently, a recognition process based on RNA (rather than specific protein binding) seems to be most parsimonious, and this suggestion has been supported by studies on the fourth chromosome of *D. melanogaster*.

The small fourth chromosome of *D. melanogaster* is considered to be entirely heterochromatic by the criteria described in Box 1, but it has 88 genes in the distal 1.2 megabases. Mapping with a *white* reporter transgene showed the presence of interspersed heterochromatic regions (inducing a variegating phenotype) (Fig. 2a) and euchromatic regions (allowing expression that results in a full red eye). Detailed examination of the region around the *Host cell factor* (*Hcf*) gene resulted in the *1360* element, which consists of remnants of a DNA transposon, being identified as a potential site for heterochromatin initiation: *D. melanogaster* with reporters lying within 10 kilobases of a *1360* element showed a variegating phenotype, indicating heterochromatin packaging and silencing, whereas *D. melanogaster* with reporters farther away from a *1360* element showed a red eye, indicating euchromatin packaging and full expression[69]. A direct test — using a *P* transposon carrying one copy of *1360*, upstream of a *white* reporter — demonstrated that *1360* contributes to silencing, because silencing of the reporter is largely lost when the adjacent *1360* is deleted. However, stable heterochromatin (resulting in a variegating phenotype) is only observed when that *P* element is located in a region close to the centromere, indicating a requirement for a high density of repeats locally and/or proximity to the pericentromeric heterochromatin, where HP1 is most abundant. Genetic analysis indicates that this silencing depends not only on HP1 and SU(VAR)3-9 but also on RNAi-pathway components, notably AUB[53]. Whether transcription occurs at the target element *1360* is unknown. Small RNA products have been recovered from *1360* and from ~40 other transposable elements in *Drosophila*[70]. It is probable that other transposable elements, in addition to *1360*, are targets for heterochromatin formation. However, it seems unlikely that all transposable-element remnants are targets, given the mapping results obtained on chromosome *4* with the *white* reporter[69]. The crucial characteristics of targets are unknown but could include the presence of start sites for transcription[53]. Many *1360* remnants contain a sequence known to function as a promoter at the multi-copy *Su(Ste)* locus, resulting in the generation of inverse transcripts that are used in the suppression of the multi-copy *Stellate* gene[71]. These results suggest that remnants of transposable elements could be targeted for silencing by a mechanism using a small RNA and that transcription of some of these elements might be involved.

Concluding remarks

Eukaryotes that tolerate large amounts of repetitious sequences in their genomes generally have both the RNAi machinery and the enzymes and structural proteins required to generate a heterochromatin structure based on H3K9 methylation. Whereas some features of the RNAi system (such as RNA-dependent RNA polymerase) and some features of the heterochromatin structure (such as DNA methylation) are used in only a subset of metazoans, this key shift in histone modification from euchromatin to heterochromatin seems to be universal (Table 1; Fig. 1). The RNAi system *per se* can limit gene expression through post-transcriptional gene silencing and can therefore eliminate some sources of damage from invading repetitious elements. However, by itself, it cannot generate the compact chromatin structures that are required to maintain chromosome integrity and chromosome function in mitosis. Hence, the suggestion that post-transcriptional gene silencing is sufficient to explain the silencing of repetitious elements seems unlikely. So, taking into account our new knowledge (described here) of the delicate balance between the need for expression and the need for silencing, an attractive model remains one in which the RNAi machinery has a key role by generating small RNAs involved in specifically targeting chromatin components (including HP1 and H3K9 methyltransferase) to silence repetitious DNA.

Although an assembly of heterochromatin structure based on binding of HP1 proteins to methylated H3K9 provides a foundation for spreading, the molecular mechanisms by which heterochromatin exerts long-range repressive effects are not fully understood. The oligomerization of chromatin-bound HP1 through the chromoshadow domains might mediate condensation. However, recent evidence suggests that HP1 binding is dynamic[72,73]. An alternative emerging view is that HP1-family proteins facilitate recruitment of regulatory proteins (effectors) that are involved in silencing and other chromosomal processes[20]. Indeed, as described earlier, HP1-family proteins mediate preferential binding of SHREC, which has HDAC activity[3]. The deacetylation of histones, which is a universal property of heterochromatic regions, might result in a lower affinity of transcription factors for target loci or could be crucial for higher-order packaging of nucleosomes, both of which would contribute to silencing. The HP1 and H3K9-methylation system might use several routes to minimize H3K9 acetylation, a key characteristic of the active state.

Evidence from different systems suggests that once triggered, a repressive chromatin structure can be sustained for many cell generations. In *S. pombe*, heterochromatin structures established by RNAi and/or DNA-binding factors are inherited *in cis* for many generations in a manner dependent on Swi6 and histone-modifying activities[7,74]. Moreover, a recent study in *Caenorhabditis elegans*, an organism known to silence repetitious DNA by using RNAi and chromatin-associated factors[75,76], showed that a single exposure to RNAi resulted in dominant silencing of a reporter gene in ~30% of the progeny for many generations[77]. A screen for mutations that affected the maintenance of silencing identified four essential genes: *hda-4* (which encodes a histone deacetylase), *K03D10.3* (which encodes a histone acetyltransferase), *isw-1* (which encodes a homologue of the chromatin-remodelling protein ISW1) and *mrg-1* (which encodes a chromodomain protein). Coupled with the observation that trichostatin A (a histone deacetylase inhibitor) relieves silencing, the results imply that maintenance of silencing is a consequence of heterochromatin formation, heritable even in the absence of the initial RNAi stimulus[77]. Although much remains to be learned about the mechanisms involved, it is clear that proper interplay of the RNAi and heterochromatin systems is crucial for the maintenance and function of our genomes. ■

Note added in proof: Two recent publications have shed light on the production of small repeat-associated RNAs in the germ line of *Drosophila*. PIWI and AUB are found associated with RNAs that are mainly antisense to transposons, whereas AGO3 is found associated with RNAs arising mainly from the sense strand. Complementary relationships between these sense and antisense RNA populations suggest that the

slicer activities of the three proteins work together to produce significant amounts of small RNA from endogenous transcripts[80,81]. Such small RNAs, which are maternally inherited, might promote both transcriptional and post-transcriptional silencing of repetitious DNA.

The demethylation of H3K4 has been suggested to be crucial for heterochromatin formation (Fig. 1), and this has now been shown in *Drosophila*[82].

In addition, a histone H2B ubiquitylation ligase complex (HULC) that facilitates Pol-II-mediated transcription of repeat elements in *S. pombe* has been identified[83]. HULC ubiquitylates H2BK119, and this, in addition to promoting euchromatic gene expression, contributes to the transcription of heterochromatic repeats.

1. Grewal, S. I. & Elgin, S. C. Heterochromatin: new possibilities for the inheritance of structure. *Curr. Opin. Genet. Dev.* **12,** 178–187 (2002).
2. Sun, F. L., Cuaycong, M. H. & Elgin, S. C. Long-range nucleosome ordering is associated with gene silencing in *Drosophila melanogaster* pericentric heterochromatin. *Mol. Cell. Biol.* **21,** 2867–2879 (2001).
3. Sugiyama, T. *et al.* SHREC, an effector complex for heterochromatic transcriptional silencing. *Cell* **128,** 491–504 (2007).
4. Wallrath, L. L. & Elgin, S. C. Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes Dev.* **9,** 1263–1277 (1995).
5. Cryderman, D. E., Tang, H., Bell, C., Gilmour, D. S. & Wallrath, L. L. Heterochromatic silencing of *Drosophila* heat shock genes acts at the level of promoter potentiation. *Nucleic Acids Res.* **27,** 3364–3370 (1999).
6. Yamada, T., Fischle, W., Sugiyama, T., Allis, C. D. & Grewal, S. I. The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. *Mol. Cell* **20,** 173–185 (2005).
7. Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297,** 2232–2237 (2002).
8. Ahmad, K. & Henikoff, S. Modulation of a transcription factor counteracts heterochromatic gene silencing in *Drosophila*. *Cell* **104,** 839–847 (2001).
9. Yasuhara, J. C. & Wakimoto, B. T. Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet.* **22,** 330–338 (2006).
10. Elgin, S. C. R. & Reuter, G. in *Epigenetics* (eds Allis, C. D., Jenuwein, T. & Reinberg, R.) 81–100 (Cold Spring Harbor Laboratory Press, Woodbury, 2007).
11. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18,** 1251–1262 (2004).
12. Locke, J., Kotarski, M. A. & Tartof, K. D. Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect. *Genetics* **120,** 181–198 (1988).
13. Eissenberg, J. C. *et al.* Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **87,** 9923–9927 (1990).
14. Shaffer, C. D. *et al.* Heterochromatin protein 2 (HP2), a partner of HP1 in *Drosophila* heterochromatin. *Proc. Natl Acad. Sci. USA* **99,** 14332–14337 (2002).
15. Reuter, G. *et al.* Dependence of position-effect variegation in *Drosophila* on dose of a gene encoding an unusual zinc-finger protein. *Nature* **344,** 219–223 (1990).
16. Tschiersch, B. *et al.* The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3-9* combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **13,** 3822–3831 (1994).
17. Greil, F., de Wit, E., Bussemaker, H. J. & van Steensel, B. HP1 controls genomic targeting of four novel heterochromatin proteins in *Drosophila*. *EMBO J.* **26,** 741–751 (2007).
18. Talbert, P. B. & Henikoff, S. Spreading of silent chromatin: inaction at a distance. *Nature Rev. Genet.* **7,** 793–803 (2006).
19. Huisinga, K. L., Brower-Toland, B. & Elgin, S. C. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115,** 110–122 (2006).
20. Grewal, S. I. & Jia, S. Heterochromatin revisited. *Nature Rev. Genet.* **8,** 35–46 (2007).
21. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406,** 593–599 (2000).
22. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292,** 110–113 (2001).
23. Jia, S., Kobayashi, R. & Grewal, S. I. Ubiquitin ligase component Cul4 associates with Clr4 histone methyltransferase to assemble heterochromatin. *Nature Cell Biol.* **7,** 1007–1013 (2005).
24. Hong, E. J. E., Villen, J., Gerace, E. L., Gygi, S. & Moazed, D. A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. *RNA Biol.* **2,** 106–111 (2005).
25. Horn, P. J., Bastie, J. N. & Peterson, C. L. A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. *Genes Dev.* **19,** 1705–1714 (2005).
26. Thon, G. *et al.* The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **171,** 1583–1595 (2005).
27. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–124 (2001).
28. Partridge, J. F., Scott, K. S., Bannister, A. J., Kouzarides, T. & Allshire, R. C. *cis*-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site. *Curr. Biol.* **12,** 1652–1660 (2002).
29. Sadaie, M., Iida, T., Urano, T. & Nakayama, J. A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. *EMBO J.* **23,** 3825–3835 (2004).
30. Cam, H. P. *et al.* Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature Genet.* **37,** 809–819 (2005).
31. Grewal, S. I. & Klar, A. J. A recombinationally repressed region between *mat2* and *mat3* loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast. *Genetics* **146,** 1221–1238 (1997).
32. Kanoh, J., Sadaie, M., Urano, T. & Ishikawa, F. Telomere binding protein Taz1 establishes Swi6 heterochromatin independently of RNAi at telomeres. *Curr. Biol.* **15,** 1808–1819 (2005).
33. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297,** 1833–1837 (2002).
34. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391,** 806–811 (1998).
35. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303,** 672–676 (2004).
36. Noma, K. *et al.* RITS acts *in cis* to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nature Genet.* **36,** 1174–1180 (2004).
37. Sugiyama, T., Cam, H., Verdel, A., Moazed, D. & Grewal, S. I. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl Acad. Sci. USA* **102,** 152–157 (2005).
38. Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119,** 789–802 (2004).
39. Irvine, D. V. *et al.* Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313,** 1134–1137 (2006).
40. Zofall, M. & Grewal, S. I. RNAi-mediated heterochromatin assembly in fission yeast. *Cold Spring Harb. Symp. Quant. Biol.* **71,** 487–496 (2006).
41. Grewal, S. I. & Moazed, D. Heterochromatin and epigenetic control of gene expression. *Science* **301,** 798–802 (2003).
42. Buhler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125,** 873–886 (2006).
43. Djupedal, I. *et al.* RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev.* **19,** 2301–2306 (2005).
44. Kato, H. *et al.* RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309,** 467–469 (2005).
45. Ayoub, N. *et al.* A novel jmjC domain protein modulates heterochromatization in fission yeast. *Mol. Cell. Biol.* **23,** 4356–4370 (2003).
46. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nature Rev. Genet.* **7,** 715–727 (2006).
47. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439,** 811–816 (2006).
48. Mochizuki, K. & Gorovsky, M. A. Small RNAs in genome rearrangement in *Tetrahymena*. *Curr. Opin. Genet. Dev.* **14,** 181–187 (2004).
49. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313,** 320–324 (2006).
50. Richards, E. J. & Elgin, S. C. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108,** 489–500 (2002).
51. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9,** 315–327 (2002).
52. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303,** 669–672 (2004).
53. Haynes, K. A., Caudy, A. A., Collins, L. & Elgin, S. C. Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Curr. Biol.* **16,** 2222–2227 (2006).
54. van Rij, R. P. *et al.* The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev.* **20,** 2985–2995 (2006).
55. Wang, X. H. *et al.* RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312,** 452–454 (2006).
56. Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* **33,** 2052–2059 (2005).
57. Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20,** 2214–2222 (2006).
58. Deshpande, G., Calhoun, G. & Schedl, P. *Drosophila* argonaute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes Dev.* **19,** 1680–1685 (2005).
59. Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J. P. & Cranston, G. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev.* **9,** 218–233 (1995).
60. Fukagawa, T. *et al.* Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nature Cell Biol.* **6,** 784–791 (2004).
61. Hall, I. M., Noma, K. & Grewal, S. I. RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl Acad. Sc.i USA* **100,** 193–198 (2003).
62. Peng, J. C. & Karpen, G. H. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nature Cell Biol.* **9,** 25–35 (2007).
63. Jia, S., Noma, K. & Grewal, S. I. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304,** 1971–1976 (2004).
64. Kim, H. S., Choi, E. S., Shin, J. A., Jang, Y. K. & Park, S. D. Regulation of Swi6/HP1-dependent heterochromatin assembly by cooperation of components of the mitogen-activated protein kinase pathway and a histone deacetylase Clr6. *J. Biol. Chem.* **279,** 42850–42859 (2004).
65. Schultz, D., Ayyanathan, K., Negorev, D., Maul, G. & Rauscher, F. R. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16,** 1855–1869 (2002).
66. Aulner, N. *et al.* The AT-hook protein D1 is essential for *Drosophila melanogaster* development and is implicated in position-effect variegation. *Mol. Cell. Biol.* **22,** 1218–1232 (2002).
67. Blattes, R. *et al.* Displacement of D1, HP1 and topoisomerase II from satellite heterochromatin by a specific polyamide. *EMBO J.* **25,** 2397–2408 (2006).

68. Huertas, D., Cortes, A., Casanova, J. & Azorin, F. *Drosophila* DDP1, a multi-KH-domain protein, contributes to centromeric silencing and chromosome segregation. *Curr. Biol.* **14,** 1611–1620 (2004).
69. Sun, F. L. *et al. cis*-Acting determinants of heterochromatin formation on *Drosophila melanogaster* chromosome four. *Mol. Cell. Biol.* **24,** 8210–8220 (2004).
70. Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5,** 337–350 (2003).
71. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11,** 1017–1027 (2001).
72. Cheutin, T. *et al.* Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **299,** 721–725 (2003).
73. Festenstein, R. *et al.* Modulation of heterochromatin protein 1 dynamics in primary mammalian cells. *Science* **299,** 719–721 (2003).
74. Nakayama, J., Klar, A. J. & Grewal, S. I. A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis. *Cell* **101,** 307–317 (2000).
75. Grishok, A., Sinskey, J. L. & Sharp, P. A. Transcriptional silencing of a transgene by RNAi in the soma of *C. elegans*. *Genes Dev.* **19,** 683–696 (2005).
76. Robert, V. J., Sijen, T., van Wolfswinkel, J. & Plasterk, R. H. Chromatin and RNAi factors protect the *C. elegans* germline against repetitive sequences. *Genes Dev.* **19,** 782–787 (2005).
77. Vastenhouw, N. L. *et al.* Gene expression: long-term gene silencing by RNAi. *Nature* **442,** 882 (2006).
78. Sullivan, B. A. & Karpen, G. H. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature Struct. Mol. Biol.* **11,** 1076–1083 (2004).
79. George, J. A. & Pardue, M. L. The promoter of the heterochromatic *Drosophila* telomeric retrotransposon, HeT-A, is active when moved into euchromatic locations. *Genetics* **163,** 625–635 (2003).
80. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128,** 1089–1103 (2007).
81. Gunawardane, L.S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315,** 1587–1590 (2007).
82. Rudolph, T. *et al.* Heterochromatin formation in *Drosophila* is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol. Cell* **26,** 103–115 (2007).
83. Zofall, M. & Grewal, S. I. HULC, a histone H2B ubiquitinating complex, modulates heterochromatin independent of histone H3 lysine 4 methylation in fission yeast. *J. Biol. Chem.* **282,** 14065–14072 (2007).

Acknowledgements We thank members of our laboratories for critical reading of the manuscript, and G. Farkas for the design of Figs 1 and 3. Our work is supported by grants from the National Institutes of Health (S.C.R.E.) and National Institutes of Health intramural support (S.I.S.G.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to S.C.R.E. (selgin@biology.wustl.edu) or S.I.S.G. (grewals@mail.nih.gov).

The complex language of chromatin regulation during transcription

Shelley L. Berger[1]

An important development in understanding the influence of chromatin on gene regulation has been the finding that DNA methylation and histone post-translational modifications lead to the recruitment of protein complexes that regulate transcription. Early interpretations of this phenomenon involved gene regulation reflecting predictive activating or repressing types of modification. However, further exploration reveals that transcription occurs against a backdrop of mixtures of complex modifications, which probably have several roles. Although such modifications were initially thought to be a simple code, a more likely model is of a sophisticated, nuanced chromatin 'language' in which different combinations of basic building blocks yield dynamic functional outcomes.

Every global traveller has experienced the disorientation of being unable to speak the local language. At present, those in the field of chromatin regulation and epigenetics are in a similar position. Like the traveller who hears words he or she does not understand, we are faced with observations that cannot be neatly categorized within previous models.

Chromatin carries numerous histone and DNA modifications, and some of these are associated with transcription. There are two proposals for the mechanistic function of these modifications. One is that chromatin packing is altered directly (either by a change in electrostatic charge or through internucleosomal contacts) to open or close the DNA polymer, thus controlling access of DNA-binding proteins such as transcription factors. The other is that the attached chemical moieties alter the nucleosome surface to promote the association of chromatin-binding proteins. Numerous discoveries made during the past decade support the idea that such modifications regulate transcription through the recruitment of effector protein complexes. Chromatin modifications are often termed 'epigenetic' marks; however, an unresolved issue in the field is the relationship between these modifications, including those established during transcription, and epigenetic inheritance (that is, the stability of these alterations during cell divisions and development). It seems that most, if not all, histone modifications are reversible, so it remains to be determined how epigenetic persistence of chromatin states is achieved, and which modifications are heritable (see pages 396 and 425).

A popular view within the field of chromatin research and in related areas, has been that DNA methylation and histone post-translational modifications (histone PTMs) correlate with either positive or negative transcriptional states. A typical model for the role of histone PTMs (Fig. 1) is that, in response to cytoplasmic signalling to transcription factors, positive-acting marks are established across promoters and open reading frames (ORFs) during gene activation through recruitment of the relevant enzymes by DNA-bound activators and RNA polymerase. In turn, negative-acting marks are laid down across genes during repression by DNA-bound repressor recruitment or across heterochromatic regions of the genome (see page 399). The resulting modifications alter the nucleosome surfaces, which then recruit regulatory protein complexes. An extension of this idea is that simple combinations of consistently behaving marks correspond to definable and predictable outcomes.

But nature may not be this simple. Recent observations suggest that, in many cases, there is not a strict division between active and repressive modification states. Instead, many of these marks have several, seemingly conflicting roles. This conflict can be explained by recent observations indicating that the function of these chromatin marks is more complex than previously thought. A prominent feature of a revised view of chromatin's role in transcription is a mechanism in which modifications establish gene activation and then reinstate repression. This dynamic behaviour leaves behind a record of activity in 'memory' marks, both positive and negative.

This review examines observations that regulation by DNA methylation or histone modification is dynamic, and that the presence of certain modifications may not indicate a unique regulatory status (that is, 'on' or 'off'). Recent developments are used to illustrate the idea that transcription is regulated by dynamic marks rather than static on/off switches. Certain histone PTMs that were previously classified as positive-acting now seem to recruit both activating and repressing effector proteins. In addition to positive-acting marks, a number of negative-acting marks are established during transcription and are localized across ORFs. It is therefore becoming difficult to classify these epigenetic mechanisms as associated with either activated or repressed genes, and a more complex, nuanced understanding is required.

Chromatin marking mechanisms in transcription

Two types of chromatin modification that regulate transcription of the protein-encoding genome are listed in Table 1. These modifications have been broadly classified into repressing and activating — in other words, they correlate with, and perhaps directly regulate, gene repression and induction. However, this binary classification is now under revision (see below).

DNA methylation occurs at cytosine residues usually within CG dinucleotides or CNG trinucleotides, and generally opposes transcription[1]. DNA methylation is localized at repeated regions and transposons in many, but not all, eukaryotes, and also occurs in genes. Promoters are normally excluded from DNA methylation because they have clusters of CpG motifs that are refractory to methylation. DNA hypermethylation is a feature of certain human cancers, causing aberrant repression of tumour suppressor genes through methylation of the CpG islands in promoters.

[1]The Wistar Institute, 3601 Spruce Street, Room 201, Philadelphia, Pennsylvania 19104, USA.

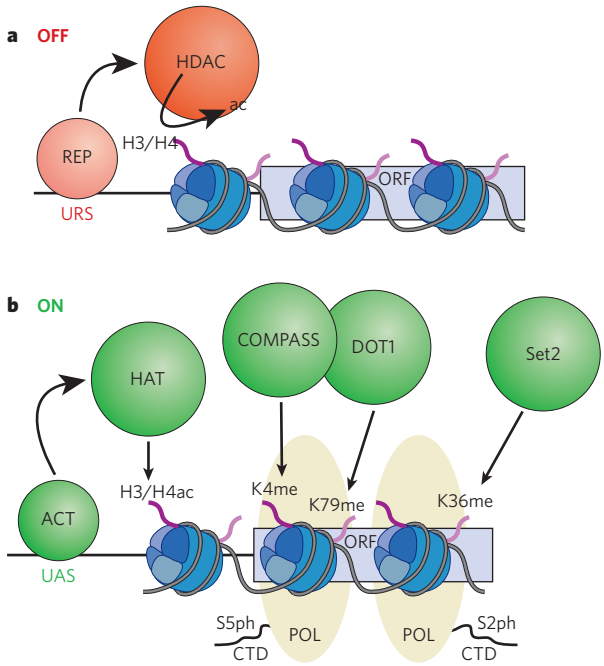


Figure 1 | Model for activation and repression. **a**, In the off state, the DNA-bound repressor (REP) at the upstream repressor site (URS) recruits negative modifiers, such as histone deacetylase (HDAC), which remove acetyl (ac) groups from histones. **b**, In the on state, DNA-bound activator (ACT) at the upstream activator site (UAS) recruits positive modifiers, such as histone acetylases (HAT), at the promoter, while DNA-bound RNA polymerase (POL) recruits histone methylases at the ORF. Early during elongation, the C-terminal domain (CTD) polymerase repeat is phosphorylated at serine 5 (S5ph), leading to recruitment of the COMPASS complex (Set1, part of the COMPASS complex, methylates H3K4) and DOT1 (which methylates H3K79). Later in elongation the CTD repeat is phosphorylated at serine 2 (S2ph), leading to recruitment of Set2 (which methylates H3K36).

There are various histone PTMs, including acetylation, phosphorylation, methylation, ubiquitylation and SUMOylation. These modifications decorate the canonical histones (H2A, H2B, H3 and H4), as well as variant histones (such as H3.1, H3.3 and HTZ.1). Most modifications localize to the amino- and carboxy-terminal histone tails, and a few localize to the histone globular domains. Because this review focuses on emerging trends relating to histone PTMs, only certain modification sites will be discussed to illustrate general concepts; there are recent reviews that provide exhaustive descriptions of the sites and modification enzymes[2].

Lysine is a key substrate residue in histone biochemistry, because it undergoes many exclusive modifications, including acetylation, methylation, ubiquitylation and SUMOylation. Acetylation and methylation involve small chemical groups, whereas ubiquitylation and SUMOylation add large moieties, two-thirds the size of the histone proteins themselves — which, owing to their bulk, may lead to more profound changes in chromatin structure. Another degree of complexity is that methylation can occur several times (mono-, di- or trimethylation) on one lysine side chain, and each level of modification can have different biological outcomes. Some of the functional outcomes of these modifications are clear. For example, there is abundant evidence that acetylation is activating, whereas SUMOylation seems to be repressing, and these two types of modification may mutually interfere. By contrast, methylation and ubiquitylation have variable effects, depending on the precise residues and contexts. For example, trimethylation of lysine 4 in histone H3 (H3K4me3) occurs at the 5′ ends of ORFs as genes become induced, whereas H3K9me3 occurs in compact pericentromeric heterochromatin, which is transcriptionally inert. Two ubiquitylation sites in the C termini of H2B and H2A correlate with active and repressed transcription, respectively.

In H3 and H4, arginine residues can also be mono- or dimethylated, and in the latter case the methyl groups can be placed symmetrically or asymmetrically on the side chain. Arginine methylation seems to be strictly activating to transcription. Serine/threonine phosphorylation is also involved in transcription. H3 phosphorylation (ph) is the most well-characterized, and H3S10ph correlates with both activated transcription and mitotic chromosome condensation — thus, with both opening and closing of chromatin — which illustrates the importance of genomic context (an idea that is developed further below). It should be noted that unbiased mass spectrometry carried out in many laboratories has led to the conclusion that the most prevalent histone PTMs in the core histones are now known. However, there may be other classes of histone PTM. For example, proline isomerization is a non-covalent alteration that occurs in H3, interconverting active and inactive conformations of the N-terminal tail for transcriptional regulation[3].

All histone PTMs are removable. Histone deacetylases (HDACs) remove acetyl groups and Ser/Thr phosphatases remove phosphate groups. Ubiquitin proteases remove mono-ubiquitin from H2B. Arginine methylation is altered by deiminases, which convert the side chain to citrulline. Two classes of lysine demethylase have recently been identified: the LSD1/BHC110 class (which removes H3K4me1 and me2)[4,5] and the jumonji class (which removes H3K4me2 and me3, H3K9me2 and me3, and H3K36me2 and me3)[6–12]. This finding put an end to heated debate about the reversibility of histone lysine methylation and whether it is the only 'true' epigenetic histone PTM.

As mentioned above, the functional consequences of histone PTMs can be direct, causing structural changes to chromatin, or indirect, acting through the recruitment of effector proteins. The structure of the H4 N-terminal tail is critical to contacts between nucleosomes. For example, acetylation of H4K16 in a naive chromatin array assembled from bacterial recombinant histones results in relaxation of the array[13]. In addition to the idea that histone PTMs directly alter nucleosome fibre structure is an expanding body of evidence that histone PTMs serve as binding surfaces for the association of effector proteins. The initial finding was for acetyl lysine, which has been shown to associate with bromodomains. In this case, acetylated H3 stabilizes binding of the histone acetyltransferase GCN5 through its bromodomain[14]. Lysine methylation provides an important switch for binding of representatives of the 'royal family' of domains, including chromodomains and tudor-domains. The key observation was methyl H3K9 association with the

Table 1 | Chromatin modifications

Mark*	Transcriptionally relevant sites†	Transcriptional role‡
DNA methylation		
Methylated cytosine (meC)	CpG islands	Repression
Histone PTMs		
Acetylated lysine (Kac)	H3 (9, 14, 18, 56), H4 (5, 8, 13, 16), H2A, H2B	Activation
Phosphorylated serine/threonine (S/Tph)	H3 (3, 10, 28), H2A, H2B	Activation
Methylated arginine (Rme)	H3 (17, 23), H4 (3)	Activation
Methylated lysine (Kme)	H3 (4, 36, 79) H3 (9, 27), H4 (20)	Activation Repression
Ubiquitylated lysine (Kub)	H2B (123§/120¶) H2A (119¶)	Activation Repression
Sumoylated lysine (Ksu)	H2B (6/7), H2A (126)	Repression
Isomerized proline (Pisom)	H3 (30–38)	Activation/repression

*The modification on either DNA or a histone.
†Well-characterized sites with regard to genomic location for DNA methylation or residues within histones for PTMs.
‡Whether the epigenetic mark is associated with activation or repression.
§Yeast (*Saccharomyces cerevisiae*).
¶Mammals.

chromodomain of heterochromatin-like protein 1 (HP1) to promote its binding to heterochromatin[15,16]. There are now other examples consistent with these initial observations.

So, simple early models suggested that certain single or combinatorial marks underlie either an active or inactive chromatin state with respect to transcription (Fig. 1). However, recent discoveries show that the functional epigenetic landscape is much more complex than previously thought.

H3K4 methylation recruits multiple effectors

The functional outcome of H3K4 methylation provides an example of the recent revision in models describing the role of modifications. As described above, H3K4 methylation has been thought to be a positive-acting histone PTM that increases as genes become active; H3K4me3 specifically localizes to the 5′ ends of ORFs. The role of this histone PTM in recruiting downstream effectors is currently under intensive investigation and, surprisingly, it has been found that a single mark recruits numerous proteins whose regulatory functions are not only activating but also repressing.

The location and timing of H3K4 methylation during gene activation led to a prediction that complexes recruited by this histone PTM are involved in transcriptional activation. Indeed, as expected, several complexes that are reported to bind H3K4me participate in gene induction (Fig. 2). The yeast SAGA complex, which contains the H3 acetyltransferase Gcn5, associates with the chromodomain protein and chromatin remodeller Chd1 (ref. 17). Chd1 binds to H3K4me, assisting in the recruitment of SAGA and leading to enhanced acetylation. CHD1 may be more important as an H3K4me binder in mammalian systems than in yeast[18]. These observations indicate that chromodomain proteins bind to methylated H3K4 as well as to methylated H3K9.

Other domains, unrelated to chromodomains, also bind to methylated H3K4. The MLL complex (also known as the Set1 K4 methylation complex) contains a WD40-repeat protein known as WDR5, which binds to H3K4me2 to contribute the trimethyl mark[19], although it is currently unclear whether WDR5 binds specifically to methylated H3K4 (refs 18, 20, 21). Another H3K4me3-interacting domain frequently found in chromatin-interacting proteins, the plant homeodomain (PHD) finger, binds preferentially to H3K4me3 to recruit positive-acting enzymatic complexes. H3K4me3 stabilizes binding of the ATP-dependent nucleosome remodelling factor (NURF) to target genes, and bromodomain PHD transcription factor (BPTF) is required for binding[22]. This pathway targets the homeodomain gene locus for activation, and the developmental effects of lowering BPTF levels are phenotypically similar to those of reducing levels of the MLL H3K4me3 enzyme, thus linking the enzyme and its binding effector protein *in vivo*. Similarly, the PHD-containing Yng1 protein in the NuA3 (nucleosomal acetyltransferase of histone H3) complex, containing the Sas3 acetyltransferase, binds preferentially to methylated H3K4 (ref. 23). Because Gcn5 and Sas3 are major H3K14 acetylation enzymes in yeast[24], it is striking that H3K4me targets both of their complexes, SAGA and NuA3 respectively, to chromatin. All of these interactions support the idea that methylated K4, a histone PTM that increases across ORFs during gene induction, serves to recruit positive-acting effector complexes.

It is therefore surprising that other complexes binding to the H3K4me3 modification are associated with transcriptional repression and not gene activation (Fig. 2). In response to DNA damage, the Sin3–Hdac1 deacetylation complex binds to H3K4me3 through the PHD domain of the Ing2 (inhibitor of growth family member 2) protein (related to Yng1 in NuA3) contained in the complex[25]. This stimulates Hdac1 activity *in vitro*, and stabilizes the recruitment of the Sin3–Hdac1 complex to target genes. This, in turn, leads to active repression of genes involved in proliferation to modulate cellular response to genotoxic damage. Another example is the JMJD2A protein, a lysine demethylase that targets H3K9me3 and H3K36me3, and is present in co-repressor complexes N-CoR (nuclear hormone co-repressor complex) and retinoblastoma. Through its tandem tudor domains, JMJD2A associates with H3K4me3 and H4K20me3 (ref. 26). Thus, it seems that H3K4me3 (and possibly H4K20me3) recruits JMJD2A to demethylate H3K9 and H3K36. Because JMJD2A associates with corepressor complexes, this recruitment presumably leads to gene repression.

These seemingly contradictory findings raise a number of questions: in particular, how can the binding of so many complexes to one histone PTM be explained? Several distinct classes of binding module recognize this mark, including the royal family domains (including chromodomains) and the PHD domains, and the different domains may bind in structurally distinct ways. The binding of single and combinations of domains to methylated lysine substrates is a complex process (see ref. 26 for an example). However, the more pressing biological question is why a single positive-acting histone PTM recruits effectors that are either activating or repressing. It may be that there is an intricate 'dance' of associations, with these changing places over time. The positive-acting complexes (such as SAGA, MLL and NuA3) may be recruited during initiation or elongation, followed by recruitment of negative-acting complexes (JMJ2D and Sin3-Hdac) during attenuation of transcription. In this case, the entire chromatin context of H3K4me3 (that is, other histone PTMs changing around it during the transcription process) would dictate the overall outcome, because other binding domains in individual complexes interact with neighbouring histone PTMs. Because both negative- and positive-acting effector complexes bind to the same histone PTM (in this case, H3K4me), these findings also strongly underscore the likely importance of DNA-bound transcription factors (repressors and activators, respectively) to discriminate during initial recruitment.

Negative marks suppress cryptic RNA initiation

In addition to H3K4 methylation across the 5′ end of genes, there are numerous other histone PTMs that occur across promoters and transcribed regions. Nucleosomes within genes constitute an obstacle to both appropriate transcription and inappropriate initiation. Thus, transcription is accompanied by disruption of histone–DNA contacts as RNA polymerase moves along, followed by reformation of nucleosomes in the wake of the enzyme. New findings show that the intricate timing of this wave is regulated by both positive- and negative-acting histone PTMs.

Positive-acting histone PTMs occur across transcribed regions and have been mechanistically linked to particular stages of transcription

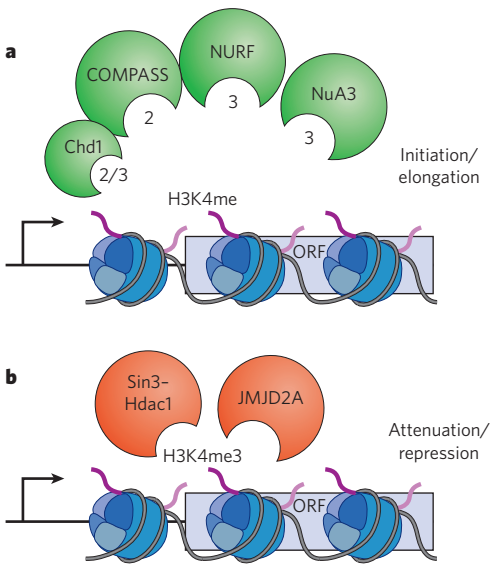


Figure 2 | H3K4me-binding proteins. a, H3K4me recruitment of activating effector proteins or complexes (Chd1, COMPASS, NURF and NuA3; numbers indicate whether they bind to H3K4me2 and/or H3K4me3) that destabilize nucleosomes during transcriptional initiation and elongation. **b**, H3K4me3 recruitment of repressing effector proteins or complexes (Sin3–Hdac1 and JMJD2A) that stabilize nucleosomes during transcription attenuation or repression.

(Fig. 3a, b). A widely accepted model is that a series of interlocking histone PTMs occur during initiation, early elongation and mature elongation, and that these events are each required for full transcriptional activation. The placement of histone PTMs is also coupled to both nucleosome clearance in the promoter and exchange of histone variants in the promoter and the ORFs, two chromatin remodelling events that occur during transcription[27–30]. The histone PTMs involve phosphorylation and acetylation of H3 through activator-mediated recruitment of kinase and acetylase complexes, followed by a complex series of ubiquitylation and methylation marks dependent on phosphorylation of RNA polymerase II. H2B becomes ubiquitylated on K123, which is, in turn, linked to methylation of H3 on K4 and K79. H2BK123 then becomes de-ubiquitylated, allowing H3K36 to be methylated[31]. However, recent evidence indicates that ORF 'marking' is more complicated than even this elaborate series of histone PTMs.

Several years ago, it was shown that transcriptional initiation can occur inside ORFs through 'cryptic' TATA boxes, and that internal initiation is detrimental to normal transcription[32]. Mutations in elongation factors that travel with RNA polymerase II cause a rise in this cryptic initiation, and increased cryptic initiation can be suppressed by mutations in RNA polymerase II. An abnormal chromatin structure is involved, manifested in altered nuclease digestion patterns correlating with the aberrant internal initiation. Furthermore, the elongation factors that suppress spurious internal start sites also facilitate nucleosome alteration during transcription[32,33].

Mechanisms that regulate cryptic initiation were recently extended to include histone PTMs and their effector proteins. Histone acetylation is involved in destabilizing or removing nucleosomes in transcribed regions[34]. One mechanism to prevent cryptic initiation is the reformation of positioned nucleosomes after the passage of RNA polymerase II, as mentioned above. Indeed, some of the histone PTMs that were previously thought to have a positive mechanistic effect on chromatin during transcription — because of their presence within ORFs and their increase during gene induction — seem instead to close the chromatin over the ORF (Fig. 3c). Thus, one role of H3K36me, which is linked to the passage of RNA polymerase II across the ORF (Fig. 1), is to recruit a deacetylase complex[35–37]. This recruitment involves a chromodomain present in the deacetylase complex, which binds to H3K36me, thereby promoting association of the histone deacetylase and removal of the nucleosome-destablizing acetyl groups. This H3K36-methylation linked deacetylation is required to suppress internal initiation.

Table 2 | Context-dependence of histone methylation

Methylated histone	Location*	Transcriptional activity†	Transcriptional status‡
H3K4me3, H3K9me3	ORFs	Inducible	Dynamic
H3K9me3, H4K20me3	Pericentromeric heterochromatin	Constitutive	Silenced
H3K4me3, H3K27me3	Embryonic stem cell bivalent domains	Constitutive	Poised for transcription

*Where each dual modification is typically found in the genome.
†Whether the dual histone PTM is detected during transcription (inducible) or is present continuously (constitutive).
‡Whether the location of the dual histone PTM marks an inducible gene (dynamic), a silenced region, or a region that is currently transcriptionally repressed but can be activated (poised).

These observations raise more questions. For example, are negative-acting epigenetic mechanisms generally required to reform nucleosomes over transcribed regions after transcription? Tantalizing clues indicate that this may be the case. First, methylated H3K9 is a well-known negative-acting histone PTM, serving to recruit HP1 through its chromodomain to lead to the formation of pericentromeric heterochromatin (see page 399). Surprisingly, H3K9me3 and HP1 have also been seen to occur within ORFs, and actually increase as transcription of genes is induced[38,39] — a result strikingly similar to the induction of methylated H3K36. This suggests that methylated H3K9 may have a similar role to methylated H3K36 in transcription — that is, to reform positioned nucleosomes over the transcribed region after the transit of RNA polymerase. Second, as described above, DNA methylation commonly occurs in promoters and genes to repress transcription. However, completion of genome-wide microarray analysis of DNA methylation in *Arabidopsis* indicates that, in addition to its expected distribution in silenced heterochromatin, DNA methylation is also common across ORFs[40,41]. Even more unanticipated is the presence of DNA methylation in the ORFs of many actively transcribed genes. So, like the histone modifications of H3K36 and H3K9 methylation, DNA methylation may help to control internal initation by reforming nucleosome structure across transcribed regions after transcription.

There is evidence of even greater complexity during the course of transcriptional elongation. As mentioned above, the transit of RNA polymerase across the transcription unit is preceded by a leading wave of initial positive histone PTMs to open the chromatin by transient displacement of nucleosomes[33]. These initial histone PTMs might later be removed, but may leave behind a 'trail' for further rounds of transcription. There are two distinct stages of transcription elongation that can be distinguished by different degrees of H3 methylation[42]. Staged evolution of histone PTMs in the ORF during the course of transcription may explain perplexing older observations in yeast that the Hos2 deacetylase is present within ORFs, its levels actually rising with gene activation, and is required for full gene activation[43]. Similarly, both H2B ubiquitylation and de-ubiquitylation is required for optimal transcription of certain genes[44]. Both the deacetylase and de-ubiquitylase may be required to remove positive-acting histone PTMs across the ORF, and set the stage for re-closure. A related possibility is that the negative-acting marks with respect to chromatin may function to set up the ORF for subsequent

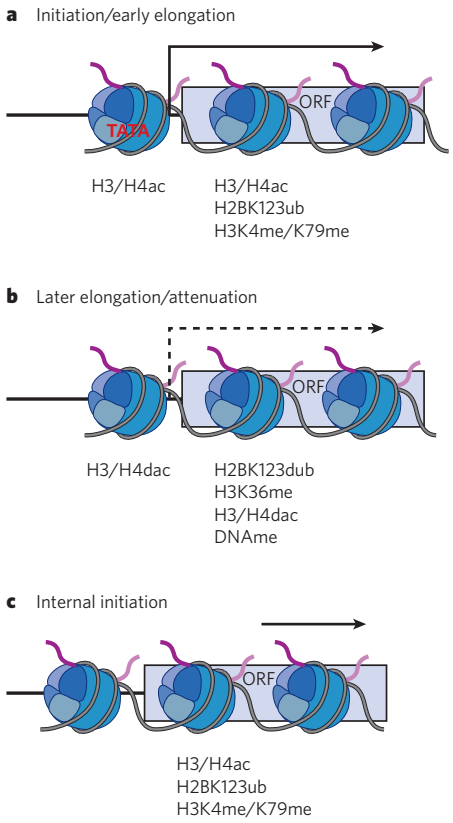


Figure 3 | Model for positive and negative chromatin marks in ORFs. **a**, Histone PTMs that occur during early transcription include H3/H4 acetylation, H2BK123 ubiquitylation and H3K4/K79 methylation. **b**, Histone PTMs that occur during late transcription include H3/H4 deacetylation (dac), H2B de-ubiquitylation (dub), H3K36 methylation and DNA methylation. Dashed arrow indicates waning transcription due to attenuation. **c**, Persistent inappropriate positive histone PTMs (H3/H4 acetylation, H2BK123 ubiquitylation and H3K4/K79 methylation) at the ORF cause internal initiation downstream of a cryptic TATA sequence.

rounds of transcription — thus, they may not only function to close the chromatin structure, but might also have a positive effect on overall transcription through stabilizing the chromatin within ORFs. So, in the face of such contradictions and subtle biological roles it seems impossible to simply classify most marks as activating or repressing .

Future challenges in deciphering histone PTM function

From these examples it is clear that chromatin marks cannot be easily interpreted — several seem to have both a positive and a negative connotation. Thus a useful analogy may be that the modifications constitute a nuanced language, in which individual marks (the 'words') become meaningful only once they are assembled and viewed within their unit array, such as a transcription unit (a 'sentence'). To put it simply, the genomic and regulatory context must be considered for the biological meaning to be understood. Because marks can indicate dynamic activity, memory of activity or poising for future activity, it is not possible to simply scan large chromosomal territories of the genome to understand a particular activity. Rather, it is necessary to observe the combination of marks and their genomic context using other signposts to interpret past history and current or future activity.

H3K4me3 and H3K9me3 are instructive examples: once thought to define activated or repressed chromatin, respectively, these marks can have vastly different messages depending on the other histone PTMs present (Table 2), and may, in fact, contradict previous interpretation (Table 1). As described above, the combination of inducible H3K4me3 and H3K9me3, detected within ORFs, indicates the current stage of dynamic transcriptional activity. By contrast, the combination of constitutive H3K9me3 and H4K20me3 seen at pericentromeric heterochromatin[45] indicates closed chromatin and repression. Furthermore, in embryonic stem cells there are unusual chromatin domains that include both H3K4me3 and H3K27me3 — again, a combination of what were thought to be positive and negative marks. These 'bivalent domains' correlate with locations of genes for transcription factors that regulate development[46]. It is thought that such genes are transcriptionally silenced but poised for activation during differentation. In neurons that differentiate from these stem cells, H3K4me3 remains constant (and H3K27me3 is lost) at genes that become active, whereas H3K27 persists (and H3K4me3 is removed) at genes that remain inactive. Thus, epigenetic status of the original stem cells can be confounding with respect to understanding transcriptional activity.

To understand this bewildering complexity, several approaches will be crucial. Elucidating the specific roles of individual histone PTMs will require multiple approaches, including temporal studies *in vivo* on synchronized populations. Chromatin immunoprecipitation analysis should reveal the order in which complexes are recruited, the concurrent chromatin modifications present, and their resulting marks at both gene and genome-wide scales during transitions from repression to activation, to attenuation, and back to repression. Interdependence of dynamic changes in the modifications must be examined by use of targeted disruption of enzymes and the modifications themselves (in the case of histone PTMs) in yeast models. The mono-, di- or trimethylation status may dictate temporal sequence of recruitment, requiring targeted disruption or knockdown of specific enzymes that set specific states. In cases in which enzymes catalyse all three methylation states (for example, Set1 within the COMPASS or MLL complexes catalysing H3K4me1, 2 and 3) then each methylation state can be inhibited using other approaches[47,48]. For example, there are specific subunits in the enzyme complexes that establish specific states, and critical binding domains in effector proteins that recognize specific states — and these can be eliminated or knocked down. These detailed experiments performed *in vivo* will help to unravel the timing of binding of specific complexes and the appearance of specific histone PTMs.

Another important approach will be *in vitro* transcription studies on chromatin arrays. Fully reconstituted chromatin templates bearing recombinant histones and DNA that have been modified in particular ways will provide the ultimate template for understanding their effect on stages of transcription. There are recent examples of *in vitro* analysis of complex chromatin transactions, including transcriptional elongation responding to histone PTMs[49] and collaboration between PTMs to accomplish optimal transcription[50]. The next step will be to reconstitute the opening and then closing events on nucleosomes during transcript elongation through the appropriate dance of histone PTMs.

Language is defined by the Webster dictionary as "A systematic means of communicating ideas… using conventionalized signs… or marks having understood meanings." This definition can be used to describe the complexity of the relationship between epigenetic marks and the biological processes they influence. As scientists, it falls to us to learn and understand this 'language', a task that we have only begun to undertake. ■

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).
2. Kouzarides, T. & Berger, S. L. in *Epigenetics* (Eds Allis, C. D., Jenuwein, T., Reinberg, D. & Caparros, M. L.) 191–209 (Cold Spring Harbor Press, New York, 2006).
3. Nelson, C. J., Santos-Rosa, H. & Kouzarides, T. Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell* **126,** 905–916 (2006).
4. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941–953 (2004).
5. Lee, M. G., Wynder, C., Cooch, N. & Shiekhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437,** 432–435 (2005).
6. Christensen, J. *et al.* RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* **128,** 1063–1076 (2007).
7. Cloos, P. A. *et al.* The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* **442,** 307–311 (2006).
8. Iwase, S. *et al.* The X-linked mental retardation gene *SMCX/JARID1C* defines a family of histone H3 lysine 4 demethylases. *Cell* **128,** 1077–1088 (2007).
9. Lee, M. G., Norman, J., Shilatifard, A. & Shiekhattar, R. Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a Polycomb-like protein. *Cell* **128,** 877–887 (2007).
10. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439,** 811–816 (2006).
11. Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125,** 467–481 (2006).
12. Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125,** 483–495 (2006).
13. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311,** 844–847 (2006).
14. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399,** 491–496 (1999).
15. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–124 (2001).
16. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410,** 116–120 (2001).
17. Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433,** 434–438 (2005).
18. Sims, R. J. & Reinberg, D. Histone H3 Lys 4 methylation: caught in a bind? *Genes Dev.* **20,** 2779–2786 (2006).
19. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121,** 859–872 (2005).
20. Couture, J. F., Collazo, E. & Trievel, R. C. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nature Struct. Mol. Biol.* **13,** 698–703 (2006).
21. Steward, M. M. *et al.* Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nature Struct. Mol. Biol.* **13,** 852–854 (2006).
22. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442,** 86–90 (2006).
23. Martin, D. G. *et al.* The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. *Mol. Cell. Biol.* **26,** 7871–7879 (2006).
24. Howe, L. *et al.* Histone H3 specific acetyltransferases are essential for cell cycle progression. *Genes Dev.* **15,** 3144–3154 (2001).
25. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442,** 96–99 (2006).
26. Huang, Y., Fang, J., Bedford, M. T., Zhang, Y. & Xu, R. M. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312,** 748–751 (2006).
27. Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11,** 1587–1598 (2003).
28. Henikoff, S. & Ahmad, K. Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* **21,** 133–153 (2005).
29. Lieb, J. D. & Clarke, N. D. Control of transcription through intragenic patterns of nucleosome composition. *Cell* **123,** 1187–1190 (2005).
30. Reinke, H. & Horz, W. Histones are first hyperacetylated and then lose contact with the activated *PHO5* promoter. *Mol. Cell* **11,** 1599–1607 (2003).
31. Shilatifard, A. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu. Rev. Biochem.* **75,** 243–269 (2006).
32. Kaplan, C. D., Laprade, L. & Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301,** 1096–1099 (2003).
33. Belotserkovskaya, R. *et al.* FACT facilitates transcription-dependent nucleosome alteration. *Science* **301,** 1090–1093 (2003).
34. Govind, C. K., Zhang, F., Qiu, H., Hofmeyer, K. & Hinnebusch, A. G. Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. *Mol. Cell* **25,** 31–42 (2007).

35. Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123,** 581–592 (2005).
36. Joshi, A. A. & Struhl, K. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell* **20,** 971–978 (2005).
37. Keogh, M. C. *et al.* Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123,** 593–605 (2005).
38. Brinkman, A.B. *et al.* Histone modification patterns associated with the human X chromosome. *EMBO Rep.* **7,** 628–634 (2006).
39. Vakoc, C. R., Mandat, S. A., Olenchock, B. A. & Blobel, G. A. Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Mol. Cell* **19,** 381–391 (2005).
40. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126,** 1189–1201 (2006).
41. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39,** 61–69 (2007).
42. Morillon, A., Karabetsou, N., Nair, A. & Mellor, J. Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. *Mol. Cell* **18,** 723–734 (2005).
43. Wang, A., Kurdistani, S. K. & Grunstein, M. Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science* **298,** 1412–1414 (2002).
44. Henry, K. W. *et al.* Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev.* **17,** 2648–2663 (2003).
45. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18,** 1251–1262 (2004).
46. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
47. Wood, A. *et al.* Ctk complex-mediated regulation of histone methylation by COMPASS. *Mol. Cell. Biol.* **27,** 709–720 (2007).
48. Xiao, T. *et al.* The RNA polymerase II kinase Ctk1 regulates positioning of a 5′ histone methylation boundary along genes. *Mol. Cell. Biol.* **27,** 721–731 (2007).
49. Pavri, R. *et al.* Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125,** 703–717 (2006).
50. Dou, Y. *et al.* Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121,** 873–885 (2005).

Acknowledgements The author thanks G. P. Moore for critical reading and valuable suggestions. Research in the author's laboratory is supported by grants from the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (berger@wistar.org).

Nuclear organization of the genome and the potential for gene regulation

Peter Fraser[1] & Wendy Bickmore[2]

Much work has been published on the *cis*-regulatory elements that affect gene function locally, as well as on the biochemistry of the transcription factors and chromatin- and histone-modifying complexes that influence gene expression. However, surprisingly little information is available about how these components are organized within the three-dimensional space of the nucleus. Technological advances are now helping to identify the spatial relationships and interactions of genes and regulatory elements in the nucleus and are revealing an unexpectedly extensive network of communication within and between chromosomes. A crucial unresolved issue is the extent to which this organization affects gene function, rather than just reflecting it.

What we know about the organization of the genome in the nucleus has been driven to a large extent by two technologies — interphase fluorescent *in situ* hybridization (FISH) and chromosome conformation capture (3C). In FISH, the relative nuclear positions of genes, genomic regions or even whole chromosomes are analysed by the hybridization of probes to nuclei fixed on glass slides. The hybridization signals are then visualized with fluorescence microscopy. In 3C, chromatin fragments in close proximity in the nucleus are captured by fixation, restriction-enzyme digestion and intramolecular ligation. The interaction between two designated genomic loci is then tested by polymerase chain reaction with primers that are specific for the loci under investigation. The limitation with both these approaches is that you only see what you are actively looking for. The advent of 4C — 3C combined with either large-scale sequencing of captured fragments or hybridization to microarrays[1–3] — has facilitated a more unbiased search for regions of the genome that interact with a particular locus, both *in cis* and *in trans*.

In this review, we discuss new data from vertebrates that help to clarify how genes function in the context of the nucleus, and that suggest that the genome's spatial organization is a key contributor to its function. We start by looking at levels of gene organization within the nuclear space occupied by individual chromosomes — the so-called chromosome territory — and ask whether some genes need to escape from this space if they are to be expressed efficiently. We then review long-range (multi-megabase) gene interactions, both *in cis* and *in trans*, and gene interactions with transcription factories. The emphasis is on the potential of such organization for gene regulation and, in particular, epigenetic regulation.

The chromosome territory as a unit of nuclear organization

The largest unit of organization of the eukaryotic genome is the chromosome. The idea that the chromosome territory is a unit of nuclear organization was advanced with the suggestion that chromosomes are discrete nuclear bodies separated by an interchromatin compartment — a more or less continuous space between adjacent chromosomes[4]. Chromosomes have preferred positions with respect to the centre or periphery of the nucleus[5] and with respect to each other[6]. Thus, a chromosome's neighbour in the nucleus is far from random — it varies between cell types, and it has consequences for a chromosome's ability to interact in *trans* with other parts of the genome, as revealed by the frequency of specific chromosome translocations[7,8]. The positional organization of chromosomes within the nucleus could therefore impinge on other aspects of genome function and, in particular, on the regulation of gene expression[9,10].

Studies have probed deep into the substructure of chromosome territories to analyse the position of sub-chromosomal regions and genes. Their results showed that for some genes, the interior, exterior or surface localization relative to their chromosome territory correlates with gene activity or inactivity[11,12]. Despite interest in interactions in the nucleus between loci on different chromosomes, one clear message that has emerged from 4C analyses is that the main sequences captured by a given locus are other regions from the same chromosome[1,3]. Thus the two-dimensional organization of the genome and the three-dimensional organization of the chromosome territory are major determinants of a gene's nuclear environment. However, experiments with 4C provide only an averaged conformation of a locus relative to its nuclear environment, which is gleaned from thousands of cell nuclei. Such a 'consensus snapshot' may not reveal transient dynamic functional states or events, especially since chromosome territories clearly do not have rigid boundaries and have a plastic organization.

Escape or eviction from the chromosome territory

Genes sometimes relocate substantial distances outside of their chromosome territory. This relocation occurs either in domains of constitutively high gene expression[13,14] or, in some instances, when gene expression is induced[15–18]. For example, the *Hoxb* gene cluster is activated at the same time as it relocates away from the chromosome territory[17]. A consequence of this relocation, as assessed by 4C, is that the 'looped out' *Hoxb* locus can now interact more with other chromosomes[1]. This finding, together with the observation of extensive intermingling of DNA from different chromosomes at the boundary of, or just outside of, chromosome territories suggests that chromosome territories might not be as discrete as previously thought[19]. It also raises the issues of why and how genes are leaving the chromosome territories and where they are going, especially since not all active genes are located outside chromosome territories[12].

Relocation outside of chromosome territories and intermingling are reduced when RNA polymerase II (Pol II) is inhibited experimentally[13,19], suggesting that these events are driven partly by the process of transcription itself. The localized chromatin decondensation that occurs

[1]Laboratory of Chromatin and Gene Expression, The Babraham Institute, Cambridge CB2 4AT, UK. [2]MRC Human Genetics Unit, Crewe Road, Edinburgh EH4 2XU, UK.

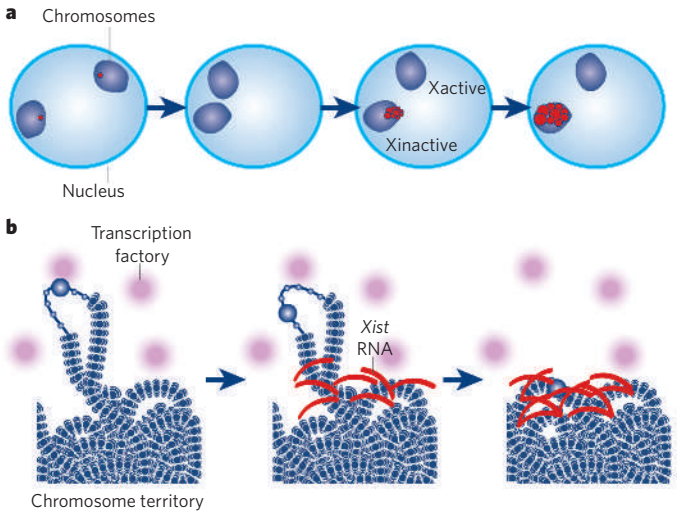


Figure 1 | Events of nuclear reorganization during X-chromosome inactivation. **a**, Soon after female embryonic stem cells start to differentiate, the two X chromosomes (purple) come together in the nucleus, and the X-inactivation centres, which initiate X-chromosome inactivation, interact[9,10]. These events occur concomitantly with the process of X-chromosome counting and choice[38] and lead to upregulation of *Xist* transcription (red) from the future inactive X chromosome (Xinactive). **b**, The coating of the inactive X chromosome by *Xist* RNA molecules excludes Pol II and the transcriptional machinery (pink) from the inactive X-chromosome territory[22]. Genes initially located outside the domain (purple circles) coated by *Xist* RNA are retracted back inside the *Xist* compartment as they become silenced through a mechanism dependent on the A repeats of *Xist* RNA[22].

in conjunction with transcriptional activation might also release constraints on chromatin's mobility and so provide it with the flexibility to facilitate looping out and to establish long-range contacts between genes and their distal regulatory elements *in cis* or *in trans*. An unresolved issue is the extent to which there might be directed chromatin movement that is driven, for example by an actin–myosin system. Compelling evidence for this idea has come from live-cell imaging of a transgene, after a transcriptional activator was targeted to it[20]. The resultant motion was curvilinear, in a direction generally perpendicular to the nuclear periphery, and perturbed by actin/myosin mutants[20]. These properties are inconsistent with the constrained-diffusion mode of chromatin motion usually seen in the mammalian genome[21], and activated endogenous genes will therefore need to be similarly analysed as they move out of their chromosome territories in living cells.

Whether the looping out is a passive response to transcriptional activity or whether it also helps to regulate this process has not been determined. A recent analysis[18] of the *Hoxd* locus in mouse development revealed that its activation along the anterior–posterior embryonic axis is accompanied by a looping out from the chromosome territory, but that activation in the limb bud, at the same stage of development, occurred in the absence of any detectable looping out from the chromosome territory in this zone even though the locus was still decondensed. This finding raises the interesting idea that the relocation of a gene outside of its chromosome territory depends on how it was activated. Evidence in favour of a role for looping out in the regulation of gene expression has come from a study of the early events in inactivation of the mammalian X chromosome. The first event that occurs after *Xist* RNA coats the chromosome to be inactivated is that Pol II is excluded from the compartment that contains the bulk of the X-chromosome territory[22]. Chromosome territories are not usually barriers to the transcriptional machinery[23,24], but the *Xist*-mediated barrier around the X chromosome seems to form at the same time as genes are silenced and as they relocate into a silent compartment (Fig. 1). The few X-linked genes that escape inactivation (that is, remain active on the inactive X chromosome) seem to be located at the edge of, or outside of, the *Xist* RNA compartment. Furthermore, if the A-repeat motif, which is required for silencing, is deleted from the *Xist* RNA, the Pol II exclusion compartment still forms, but genes usually scheduled for inactivation are not silenced or modified epigenetically and do not relocate to the inside of the *Xist* compartment. These results link the physical relocation of X-linked genes with the silencing mechanism and suggest that the silencing depends on the architecture of the nucleus and chromosome[22]. Although the strict order of events is not known, an intriguing possibility is that the A repeats, and the factors that interact with them, capture X-linked genes and consign them to the interior, thus leading eventually to epigenetic silencing.

Colocalization of active genes in the nuclear space

Are genes located at the periphery of, or outside of, chromosome territories just moving randomly in a different nuclear environment from those restricted to the chromosome interior, or are they interacting with, or tethered to, discrete sites? Actively transcribed genes localize at focal concentrations of Pol II called 'transcription factories'. The number of discrete factories visible in the nucleus seems to be lower than the number of expressed genes, suggesting that multiple genes share the same factory (Fig. 2). Indeed, both RNA FISH and 3C confirm the nuclear colocalization of active alleles of genes that are separated by tens of megabases in *cis*, or even of those located on different chromosomes[25]. The gene and factory may require a physical association for efficient transcription, since actively transcribed alleles are almost always found in the factories[25,26]. However, the process is not as simple as getting to a factory, becoming active and then staying there. Studies done in live cells have shown that individual genes are transcribed in bursts or pulses of production[27]. FISH studies support this transcriptional pulsing model and have shown that temporarily quiescent alleles of 'active' genes are located away from factories, suggesting that gene mobility is an important factor in gene control[25].

However, related studies have led to a different interpretation of the spatial colocalization of actively transcribed genes. The zone of spatial colocalization, or juxtaposition, of active gene loci was thought to be larger than the predicted size of transcription factories as measured by foci of bromodeoxyuridine UTP incorporation[28]. Rather, researchers have suggested that the active genes congregate at, or around, splicing-factor-enriched speckles — accumulations of messenger-RNA-splicing factors that have a diameter of 0.5-3.0 μm (Fig. 2)[14,29]. Some active genes are found at the speckles' boundaries, and specific pairs of genes in *cis* and *trans* have been shown to co-associate with the same speckle[14,29,30]. However, transcription sites are clearly distinct from splicing-factor-enriched speckles — nascent RNA can be detected at the edges, but most transcription occurs outside the speckles[31]. One suggestion is that some pre-mRNAs transit through the speckles as part of their processing and export[32,33].

The idea of nuclear zones in which very active genes are colocalized is compatible with the results from 4C analysis of the β-globin gene (*Hbb*) in fetal liver, and the ubiquitously expressed *Rad23a* gene in both liver and brain. In both cases, the active genes were associated with several clusters of other active genes. Each cluster was 150–200 kilobases in size and located mainly within a 70 megabase region in *cis*. But associations with gene clusters were also detected *in trans*[3] (Table 1). The 'interactions' were not restricted to a specific gene within each cluster nor to specific parts of genes such as promoters, and they were far too numerous to suggest that they all occurred simultaneously in a single cell. These findings raise the issue of whether the interactions detected by 4C originate from a probabilistic three-dimensional chromosome structure that differs from cell to cell or from multiple dynamic interactions that occur transiently in all or most cells. The multiple-interactions explanation is consistent with a study that used FISH to look at recognizable folding states of a 4.3 megabase region from mouse chromosome 14 that contains alternating gene-dense regions and gene deserts. FISH signals from the gene-dense regions were juxtaposed spatially in about 20% of the nuclei[34], and nuclear colocalization seemed not to be restricted to active gene regions.

Nuclear colocalization of genes and regulatory elements

In tissue in which *Hbb* is not expressed (brain), 4C showed that the interactions captured with *Hbb* were with other silent gene clusters, including clusters of olfactory-receptor genes that are *in cis* with, but distant from, *Hbb*[3] (Table 1). Although there is no evidence that in this instance the associations contribute functionally to the silent state of these gene clusters, the spatial organization of olfactory-receptor genes does seem to be involved in their extraordinary regulation in olfactory sensory neurons. Only 1 of the 1,300 olfactory-receptor genes is expressed in a given neuron, and then only from 1 allele of that gene. An enhancer element (*H*) required for expression of olfactory-receptor genes has been identified[35]. A quarter of sequences identified as interacting with this element by 4C analysis were immediately upstream of olfactory-receptor genes[36]. Like the 4C interactions described for *Hbb*, most (75%) of the olfactory-receptor genes captured with *H* were located *in cis*, although some interactions were *in trans* (Table 1). This predominance broadly reflects the relative frequency with which different olfactory-receptor genes are expressed and suggests that expression of an olfactory-receptor gene depends on its interaction with *H* in the nucleus (Fig. 3). Consistent with this idea, DNA FISH showed that *H* colocalizes with an allele of a specific olfactory-receptor gene in about 30% of the cells that express that gene. Since only one allele of *H* is thought to be active in this respect (the other is methylated, unusually at CpA sequences), expression of olfactory-receptor genes would be restricted to the allele that is in contact with *H*. Indeed, combined RNA and DNA FISH for *H* DNA and a specific olfactory-receptor RNA in *trans* showed that *H* is in contact with the actively transcribed allele in 85% of cells with a FISH signal for olfactory-receptor RNA. Taken together, these results suggest that *H* is in contact with the active olfactory-receptor gene specifically during a transcriptional pulse but may dissociate during periods of transcriptional inactivity.

Transient interactions between regulatory elements and genes in the nucleus have also been associated with coordinately regulated gene expression in other examples. Interchromosomal interactions between the interferon gamma (*Ifng*) gene and the locus control region of the T-helper 2 (T_H2) cytokine locus are found by 3C and FISH in the nuclei of naive $CD4^+$ T cells[37] (Table 1). This interaction is thought to hold the *Ifng* and T_H2 cytokine gene loci in a poised state that can respond rapidly to T-cell activation by expression of both gene loci, but only at very low levels. Later, after the decision to differentiate into T_H1 or T_H2 cells has been made, and expression of *Ifng* or T_H2 cytokines is very high, these interchromosomal associations are lost in favour of intrachromosomal ones. It will now be interesting to analyse the interaction of these loci with other regions of the genome during T-cell activation and differentiation by use of 4C approaches.

These studies of olfactory-receptor genes and cytokine genes clearly implicate nuclear interactions in gene-expression states that are then epigenetically stable, and may even implicate nuclear organization in epigenetic 'decisions'. The paradigms for investigation of epigenetic mechanisms that underpin allelic choice are X-chromosome inactivation and imprinting. Strikingly, nuclear interactions have been implicated in both. The two X chromosomes transiently colocalize, and the X-inactivation centres interact physically during the differentiation of female embryonic stem cells[9,10]. The timing of this nuclear interaction is concurrent with the onset of X-chromosome inactivation[38] (Fig. 1). It is also within the time frame that would implicate it in the intriguing process whereby the number of X chromosomes per nucleus is 'counted' and the 'choice' is made to inactive one of them.

The other system in which transient nuclear interaction happens on homologous chromosomes is imprinting[39]. In mouse neonatal liver, a 4C approach showed that the imprinting control region (ICR) of *H19* had many *trans* interactions in addition to its abundant *cis* interactions[2]. Importantly, sequences from multiple different chromosomes could be captured within the same 4C clone after intramolecular ligation, suggesting that the nuclear associations were not just pairwise, but that multiple genomic loci could co-associate simultaneously with the *H19* ICR in a single nucleus. In mice in which the maternally derived *H19* alleles could be distinguished from the paternally derived ones (*Mus musculus* versus *Mus spretus* origins), about 75% of the sequences captured were specific for the maternally derived allele. Hence, the epigenetic status of the *H19* ICR, which is methylated on the paternally derived allele, determines the patterns of most of its intra- and interchromosomal interactions. As well as maternal-specific interactions *in cis* with another imprinted domain on mouse chromosome 7, interactions were also detected with known, or predicted, imprinted regions on other chromosomes (Table 1). 3C analysis confirmed that the *H19* ICR interacted with a differentially methylated region on chromosome 18 that is implicated in the control of imprinting there. Hence, the nuclear interaction might not one be of gene with gene, or gene with regulatory element, but of two regulatory elements with each other. The functional significance of this finding was demonstrated by the altered expression of two genes, *Osbpl1a* and *Impact*, from within the chromosome 18 imprinted domain in mice

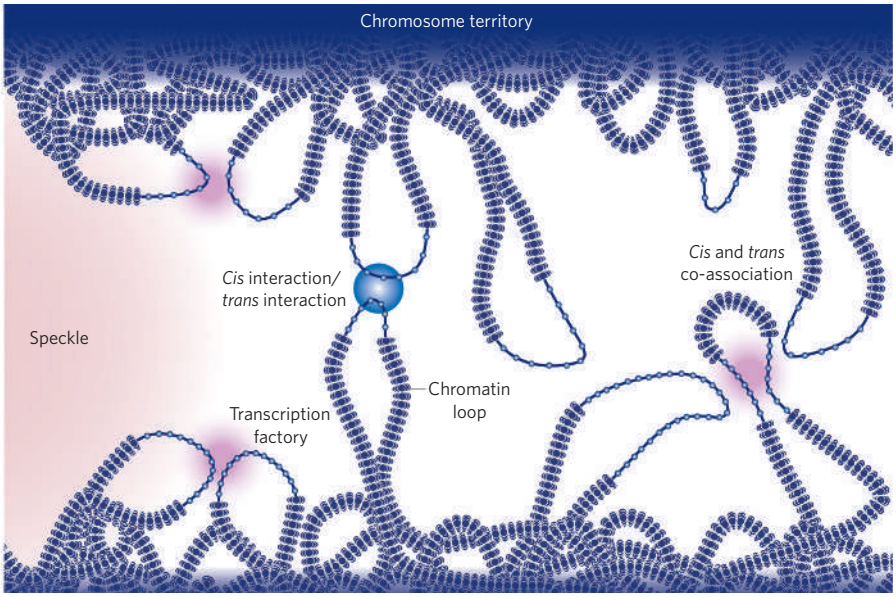


Figure 2 | Colocalization of genes in the nucleus for expression or coregulation. Active genes on decondensed chromatin loops that extend outside chromosome territories can colocalize both *in cis* and *in trans* at sites in the nucleus with local concentrations of Pol II (namely transcription factories; dark pink) and adjacent to splicing-factor-enriched speckles (pale pink). Interactions can also occur between regulatory elements and/or gene loci and lead to coregulation in *trans* (blue circle).

Table 1 | Long-range intrachromosomal and interchromosomal interactions detected by 3C and 4C

Starting locus	Captured interacting locus	Intra- or interchromosomal	Cell type	Reference
3C				
Hbb (active)	Erythroid associated factor (*Eraf*), uroporphyrinogen III synthase (*Uros*)	Intra	Fetal liver	25
T_H2 locus control region	*Ifng*	Inter	Naive $CD4^+$ T cells	37
Tsix (*Xic*)	Other Xinactive-specific transcript (*Tsix*) allele	Inter	Embryonic stem cells	9
H19 imprinting control region (maternal allele)	*Osbpl1a*, *Impact*	Inter	Neonatal liver	2
4C				
Rad23a (active)	Multiple active gene regions	Intra > inter	Fetal liver and brain	3
Hbb (active)	Multiple active gene regions including those containing *Eraf*, *Uros* and *Kcnq1*	Intra > inter	Fetal liver	3
Hbb (inactive)	Silent gene clusters: for example, olfactory-receptor genes	Intra	Fetal brain	3
H enhancer	Expressed olfactory-receptor allele	Intra > inter	Olfactory sensory neurons	36
H19 imprinting control region	*Wsb1*, *Nf1*	Intra-	Fibroblasts	40

with a mutated maternal *H19* ICR. The multiple (>100) nuclear interactions of *H19* ICR in this study[2] contrasted sharply with the results of a superficially similar experiment showing only one interaction *in cis* and two *in trans* for the maternal allele of the *H19* ICR[40]. Why such different pictures of the breadth of *H19* ICR interactions arise from these two experiments is unclear, but they might, for example, represent differences in the detailed experimental protocols, or in the cells and cell lines used in each case — liver and differentiating embryonic stem cells versus a fibroblast cell line.

Future directions and perspectives

The overall picture that is emerging, with the aid of rapid technological advances, is that genes have some independence in the nucleus but do not necessarily function in isolation from each other. They are taken to, or find themselves in, different nuclear environments, which are often shared by other genes. Although many genes may relocate for specific purposes in accordance with their own activation status, other genes nearby on the DNA sequence may be taken along for the ride. The constraints imposed by the functional organization of the nucleus could provide the selective pressure to maintain clusters of broadly expressed, apparently functionally unrelated, genes together on chromosomes during evolution[41,42]. In particular, multiple active genes and gene clusters are often located together at places in the nucleus that have high local concentrations of transcriptional and mRNA-processing machinery. A study of the spatial relationship of genes both with Pol II and with splicing-factor-enriched speckles will help to determine whether these nuclear zones are organized mainly by the preference for specific genes to be transcribed together (transcription factories) or to ensure efficient transcription and co-transcriptional mRNA processing (mRNA expression factories). An exciting new dimension to this area is added by the detection of interactions in the nucleus between regulatory elements. The colocalization of these elements is probably not because they are both recruited to a particular nuclear compartment, but because direct protein–protein interactions between the loci promote or restrain the genes' ability either to sample other nuclear environments necessary for their activation or repression or to directly modify the chromatin structure of the interacting alleles. To differentiate between these models we need to know more about the proteins that mediate the interactions — especially those that occur *in trans*. One clue that we have about this is that the interactions of the maternal *H19* ICR are dependent on the binding of the protein CTCF[2,40].

Two extreme camps have emerged about the functional relevance of nuclear organization. One camp thinks that nuclear organization merely reflects the functional processes occurring in the genome (for example, transcription). The other thinks that nuclear organization is a major factor in regulating the genome's function. The data we have reviewed here suggest that the dynamic spatial organization of the nucleus both reflects and shapes genome function[42]. There are also exciting examples of epigenetic 'decisions' that are made or influenced by *trans* interactions between specific loci, and these will continue to provide attractive model systems that further our understanding of the mechanistic importance of these levels of nuclear organization. We now have a picture of a genome that is 'structured', not in a rigid three-dimensional network, but in a dynamic organization with preferred or probabilistic conformations experienced by similarly programmed cells. This organization clearly changes during normal development and differentiation, and is itself 're-programmable' — for example, during the process of cloning by nuclear transfer[43,44]. The 'tough nut' of the nucleus is starting to be cracked as more complex molecular events are related to its three-dimensional functional organization. The studies reviewed here reveal that the scope of long-range *cis* and *trans* functional interactions is much greater than imagined, and there is no reason to think that this is the limit. Mobility is an important factor in genome function and could be a control point for several nuclear events. Thus, analysis in living cells will have a major role in future experiments. Exciting times lie ahead

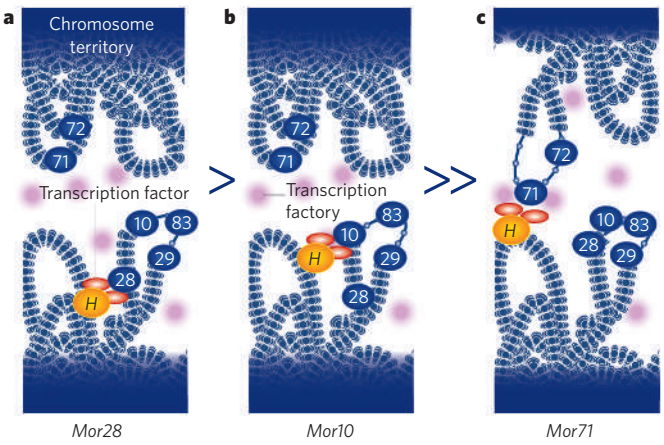


Figure 3 | *Cis* and *trans* interactions of the *H* enhancer and olfactory-receptor genes. A quarter of 4C products captured in the nuclei of sensory neurons with the olfactory-receptor *H* enhancer (orange) are olfactory-receptor genes themselves (numbered circles)[36]. More than half of these are the promoter of *Mor28*, which is adjacent to *H* on mouse chromosome 14 (**a**). The next most frequent interaction found was with the promoter of *Mor10*, which is also in *cis* with *H* (**b**). Much less frequent were interactions with olfactory-receptor genes (for example, *Mor71*) located on other mouse chromosomes, and so in *trans* with *H* (**c**). Each olfactory neuron expresses only one allele of one olfactory-receptor gene, and the frequencies of 4C interactions with *H* reflect the relative frequencies of expression of different olfactory-receptor genes in the population of sensory neurons. The *H* enhancer might function by recruiting transcription factors (red) and high concentrations of Pol II (present in transcription factories) to the expressed gene.

as the wealth of biochemical information known about specific nuclear processes is mapped onto the global coordinates of the nucleus to create a more holistic view of functional organization of the genome and its role in delivering gene-expression programmes. ■

1. Wurtele, H. & Chartrand, P. Genome-wide scanning of *HoxB1*-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res.* **14,** 477–495 (2006).
2. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet.* **38,** 1341–1347 (2006).
3. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38,** 1348–1354 (2006).
4. Cremer, T. *et al.* Chromosome territories — a functional nuclear landscape. *Curr. Opin. Cell Biol.* **18,** 307–316 (2006).
5. Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* [online] **3,** e157 (2005) (doi: 10.1371/journal.pbio.0030157).
6. Parada, L. & Misteli, T. Chromosome positioning in the interphase nucleus. *Trends Cell Biol.* **12,** 425–432 (2002).
7. Bickmore, W. A. & Teague, P. Influences of chromosome size, gene density and nuclear position on the frequency of constitutional translocations in the human population. *Chromosome Res.* **10,** 707–715 (2002).
8. Parada, L. A., McQueen, P. G. & Misteli, T. Tissue-specific spatial organization of genomes. *Genome Biol.* [online] **5,** R44 (2004) (doi: 0.1186/gb-2004-5-7-r44).
9. Xu, N., Tsai, C. L. & Lee, J. T. Transient homologous chromosome pairing marks the onset of X inactivation. *Science* **311,** 1149–1152 (2006).
10. Bacher, C. P. *et al.* Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nature Cell Biol.* **8,** 293–299 (2006).
11. Kurz, A. *et al.* Active and inactive genes localize preferentially in the periphery of chromosome territories. *J. Cell Biol.* **135,** 1195–1205 (1996).
12. Mahy, N. L., Perry, P. E., Gilchrist, S., Baldock, R. A. & Bickmore, W. A. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J. Cell Biol.* **157,** 579–589 (2002).
13. Mahy, N. L., Perry, P. E. & Bickmore, W. A. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol.* **159,** 753–763 (2002).
14. Brown, J. M. *et al.* Coregulated human globin genes are frequently in spatial proximity when active. *J. Cell Biol.* **172,** 177–187 (2006).
15. Volpi, E. V. *et al.* Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* **113,** 1565–1576 (2000).
16. Williams, R. R., Broad, S., Sheer, D. & Ragoussis, J. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp. Cell Res.* **272,** 163–175 (2002).
17. Chambeyron, S. & Bickmore, W. A. Chromatin decondensation and nuclear reorganization of the *HoxB* locus upon induction of transcription. *Genes Dev.* **18,** 1119–1130 (2004).
18. Morey, C., Da Silva, N. R., Perry, P. & Bickmore, W. A. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to *Hox* gene activation. *Development* **134,** 909–919 (2007).
19. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* [online] **4,** e138 (2006) (doi: 10.1371/journal.pbio.0040138).
20. Chuang, C. H. *et al.* Long-range directional movement of an interphase chromosome site. *Curr. Biol.* **16,** 825–831 (2006).
21. Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12,** 439–445 (2002).
22. Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for *Xist* RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **20,** 2223–2237 (2006).
23. Abranches, R., Beven, A. F., Aragon-Alcaide, L. & Shaw, P. J. Transcription sites are not correlated with chromosome territories in wheat nuclei. *J. Cell Biol.* **143,** 5–12 (1998).
24. Sadoni, N. & Zink, D. Nascent RNA synthesis in the context of chromatin architecture. *Chromosome Res.* **12,** 439–451 (2004).
25. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genet.* **36,** 1065–1071 (2004).
26. Ragoczy, T., Bender, M. A., Telling, A., Byron, R. & Groudine, M. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.* **20,** 1447–1457 (2006).
27. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16,** 1018–1025 (2006).
28. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* **109,** 1427–1436 (1996).
29. Moen, P. T. Jr *et al.* Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol. Biol. Cell* **15,** 197–206 (2004).
30. Shopland, L. S., Johnson, C. V., Byron, M., McNeil, J. & Lawrence, J. B. Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J. Cell Biol.* **162,** 981–990 (2003).
31. Xie, S. Q., Martin, S., Guillot, P. V., Bentley, D. L. & Pombo, A. Splicing speckles are not reservoirs of RNA polymerase II, but contain an inactive form, phosphorylated on serine2 residues of the C-terminal domain. *Mol. Biol. Cell* **17,** 1723–1733 (2006).
32. Johnson, C. *et al.* Tracking *COL1A1* RNA in osteogenesis imperfecta. Splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *J. Cell Biol.* **150,** 417–432 (2000).
33. Molenaar, C., Abdulle, A., Gena, A., Tanke, H. J. & Dirks, R. W. Poly(A)+ RNAs roam the cell nucleus and pass through speckle domains in transcriptionally active and inactive cells. *J. Cell Biol.* **165,** 191–202 (2004).
34. Shopland, L. S. *et al.* Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. *J. Cell Biol.* **174,** 27–38 (2006).
35. Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor–one olfactory neuron rule in mouse. *Science* **302,** 2088–2094 (2003).
36. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126,** 403–413 (2006).
37. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435,** 637–645 (2005).
38. Morey, C. & Bickmore, W. Sealed with a X. *Nature Cell Biol.* **8,** 207–209 (2006).
39. LaSalle, J. M. & Lalande, M. Homologous association of oppositely imprinted chromosomal domains. *Science* **272,** 725–728 (1996).
40. Ling, J. Q. *et al.* CTCF mediates interchromosomal colocalization between *Igf2/H19* and *Wsb1/Nf1*. *Science* **312,** 269–272 (2006).
41. Sproul, D., Gilbert, N. & Bickmore, W. A. The role of chromatin structure in regulating the expression of clustered genes. *Nature Rev. Genet.* **6,** 775–781 (2005).
42. Chakalova, L., Debrand, E., Mitchell, J. A., Osborne, C. S. & Fraser, P. Replication and transcription: shaping the landscape of the genome. *Nature Rev. Genet.* **6,** 669–677 (2005).
43. Eggan, K. *et al.* Mice cloned from olfactory sensory neurons. *Nature* **428,** 44–49 (2004).
44. Li, J., Ishii, T., Feinstein, P. & Mombaerts, P. Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature* **428,** 393–399 (2004).

Acknowledgements P.F. is a Senior Fellow of the Medical Research Council UK and receives support from the Biotechnology and Biological Sciences Research Council, UK. W.B. is a Centennial fellow of the James S. McDonnell Foundation, is supported by the Medical Research Council UK and acknowledges the contribution of the EU FP6 Epigenome Network of Excellence.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to W.B. (w.bickmore@hgu.mrc.ac.uk) or P.F. (peter.fraser@bbsrc.ac.uk).

NATURE|Vol 447|24 May 2007|doi:10.1038/nature05917

Epigenetic inheritance in plants

Ian R. Henderson[1] & Steven E. Jacobsen[1]

The function of plant genomes depends on chromatin marks such as the methylation of DNA and the post-translational modification of histones. Techniques for studying model plants such as *Arabidopsis thaliana* have enabled researchers to begin to uncover the pathways that establish and maintain chromatin modifications, and genomic studies are allowing the mapping of modifications such as DNA methylation on a genome-wide scale. Small RNAs seem to be important in determining the distribution of chromatin modifications, and RNA might also underlie the complex epigenetic interactions that occur between homologous sequences. Plants use these epigenetic silencing mechanisms extensively to control development and parent-of-origin imprinted gene expression.

Eukaryotic genomes are covalently modified with a diverse set of chromatin marks, which are present on both the DNA and the associated histones (see page 407). Although these changes do not alter the primary DNA sequence, they are frequently heritable through cell division, sometimes for multiple generations, and can thus often be classified as epigenetic marks. These conserved epigenetic marks have been found to influence many aspects of gene expression and chromosome biology, and they have characteristic genomic distributions.

The size of eukaryotic genomes varies extensively and does not correlate with gene number[1]. This is often because of the presence of large amounts of non-gene sequences, which can include pseudogenes, transposable elements, integrated viruses and simple repeats[1]. At the chromosomal level, genomes are organized into euchromatin, which is gene-rich, and heterochromatin, which is repeat-rich[2]. Heterochromatin is defined by three main properties: greater compaction than other genomic regions during interphase, lower accessibility than other regions to transcription and recombination machinery, and the formation of structured nucleosome arrays[2] (see page 399). The defining characteristics of heterochromatin depend on epigenetic information, including post-translational modification of histones and methylation of cytosine bases in DNA[2,3]. The silencing of transposable-element sequences within heterochromatin is probably a genome-defence strategy. However, heterochromatin can also have important roles during chromosomal segregation[4], and transposons and epigenetic silencing have been shown to both modulate gene expression and contribute to *cis*-regulatory sequences[5,6]. Plant systems have been a rich source for the study of epigenetic inheritance, and examples of important discoveries include transposable elements[7], paramutation[8], small interfering RNAs (siRNAs)[9] and RNA-directed DNA methylation[10].

Genomic resources for studying the model plant *Arabidopsis thaliana* have begun to provide insight into the epigenetic 'landscape' of this organism[11,12]. *A. thaliana* has a compact ~130-megabase (Mb) genome, although it contains considerable amounts of heterochromatin, which is repeat-rich and largely located in the centromeric and pericentromeric regions[13,14] (Fig. 1). High-resolution mapping of cytosine methylation by using whole-genome microarrays has confirmed previous reports, showing that this modification co-localizes with repeat sequences and with the centromeric regions[11,12,15]. Fewer than 5% of expressed genes were shown to have methylated promoters, although about one-third of genes were methylated in their open reading frame[11,12]. The significance of methylation in the body of a gene is not fully understood, but such methylation was found to correlate with genes that are both highly transcribed and constitutively expressed[11,12]. By contrast, genes with methylated promoters had lower expression levels and frequently had tissue-specific expression patterns[11,12]. This distribution of cytosine methylation is in contrast to that observed in mammalian genomes, which are often densely methylated but have hypomethylated CG islands in gene promoters[3]. It will be important to describe the 'methylome' of other repeat-rich plant genomes, such as those of the grasses, to test the generality of the patterns observed in *A. thaliana*. Here, we review the emerging and prominent role of RNA in epigenetic inheritance in plants and how such mechanisms are used to control development.

Mediating silencing with RNA

A central question in understanding the epigenetic regulation of genomes is how sequences are recognized or avoided as targets for silencing. There is an increasing appreciation that siRNAs, which are generated by the RNA interference (RNAi) pathway, can provide sequence specificity to guide epigenetic modifications in a diverse range of eukaryotes. Well-studied examples include transcriptional silencing in yeast[16] (see page 399), cytosine methylation in plants[10,17] and genome rearrangements in ciliates[18]. RNA-directed DNA methylation was discovered in tobacco, in which genomic sequences homologous to infectious RNA viroids were found to become cytosine methylated[10]. Subsequently, the expression of double-stranded RNA (dsRNA) in plants was shown to generate siRNAs and cause dense cytosine methylation of homologous DNA in all sequence contexts[19]. This is reflected by the high coincidence of endogenous siRNA clusters with methylated sequences and repeats in *A. thaliana*[11,12,15,20].

All known *de novo* DNA methylation in *A. thaliana* is carried out by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), which is a homologue of the mammalian DNA methyltransferase 3 (DNMT3) enzymes[21–24] (Fig. 2b). DRM2 can be targeted to a sequence by siRNAs generated from the expression of either direct or inverted repeats[23,24]. Plants encode multiple homologues of the RNAi-machinery components, some of which are specialized for function in RNA-directed DNA methylation[25,26]. The endoribonuclease DICER-LIKE 3 (DCL3) generates 24-nucleotide siRNAs, which are loaded into the PAZ- and PIWI-domain-containing protein ARGONAUTE 4 (AGO4)[26–31] (Fig. 2a). These AGO4-associated siRNAs are proposed to guide the cytosine-methyltransferase activity of DRM2 (refs 26–31). The mechanism by which siRNAs target epigenetic modifications is poorly understood and could involve either DNA–RNA or RNA–RNA hybridization events. Interestingly, epigenetic modifications guided by AGO4 in

[1]Department of Molecular, Cell and Developmental Biology, Howard Hughes Medical Institute, University of California, Los Angeles, California 90095, USA.

A. thaliana have been shown to depend partly on the RNaseH ('slicer') catalytic activity of AGO4 (ref. 30). This could be taken as support for RNA–RNA hybridization having an important role in the targeting of epigenetic modifications.

The accumulation of siRNAs associated with RNA-directed DNA methylation in *A. thaliana* often depends on RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and the plant-specific protein NUCLEAR RNA POLYMERASE IV A (also known as NUCLEAR RNA POLYMERASE D 1A; NRPD1A), which are involved in a putative amplification pathway[26,32–35] (Fig. 2a). Together, RDR2 and NRPD1A might generate dsRNA substrates for DCL3 to process into siRNAs, although how these proteins are recruited to target loci is unknown. Several loci also show dependence on AGO4 and DRM2 for siRNA accumulation, suggesting that there might be a feedback loop between transcriptional silencing and siRNA generation[24,26].

NRPD1A functions in a complex with NRPD2. A variant of this NRPD complex, which contains NRPD1B instead of NRPD1A, is also required for RNA-directed DNA methylation but participates less frequently in siRNA accumulation[33,35] (Fig. 2a). One possible function for the NRPD1B-containing complex is to generate a target transcript that can hybridize with siRNA-loaded AGO4-containing complexes. Indeed, AGO4 has been observed to bind directly to NRPD1B[28]. The SWI–SNF-family chromatin-remodelling protein DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1) is also required for RNA-directed DNA methylation and could function to facilitate access of DRM2 to target DNA[27,36]. Recently, several proteins in the RNA-directed DNA-methylation pathway have been found to localize to distinct nuclear bodies, including the Cajal body, which is a centre for the processing and modification of many non-coding RNAs[28,29]. Localization to these bodies might be required for the efficient loading of AGO4-containing complexes with siRNA before these complexes travel to the nucleoplasm and, together with DRM2, direct RNA-directed DNA methylation.

Plants show extensive methylation of cytosine bases in the CG, CNG (where N denotes any nucleotide) and CHH (where H denotes A, C or T) sequence contexts[37]. By contrast, most cytosine methylation in mammals is found in the CG sequence context[3,38]. CG methylation is maintained by the homologous proteins METHYLTRANSFERASE 1 (MET1) and DNMT1 in plants and mammals, respectively[39,40] (Fig. 2b). DNMT1 has a catalytic preference for hemimethylated substrates, providing an attractive model for the efficient maintenance of CG methylation after DNA replication and during cell division[38]. Most non-CG methylation in plants is maintained redundantly by DRM2 and the plant-specific protein CHROMOMETHYLASE 3 (CMT3)[23,37] (Fig. 2b); however, some loci show residual non-CG methylation in *drm1 drm2 cmt3* triple mutants, which might be maintained by MET1 (ref. 25). Non-CG methylation differs from CG methylation, because it seems to require an active maintenance signal after DNA replication. At some loci, siRNAs seem to provide this signal, acting through DRM2 activity: for example, at the *MEA-ISR* locus (*MEDEA INTERSTITIAL SUBTELOMERIC REPEATS* locus, an array of seven tandem repeats located downstream of the *MEDEA* gene), the repeats lose all non-CG methylation in *drm2* mutants and in several RNAi-pathway mutants such as *ago4* and *rdr2* (refs 23, 37). By contrast, other loci — for example, the SINE-class retrotransposon *AtSN1* — completely lose non-CG methylation only in *drm1 drm2 cmt3* triple mutants. At *AtSN1*, CMT3 contributes to the maintenance of both CNG methylation and asymmetrical (CHH) methylation. The activity of CMT3 largely depends on the main methyltransferase for H3K9 (the lysine residue at position 9 of histone H3) — SU(VAR)3-9 HOMOLOGUE 4 (SUVH4; also known as KRYPTONITE) — showing that histone methylation is also an important signal for the maintenance of non-CG methylation[41,42]. At present, the factors that determine the relative importance of the RNAi pathway and histone methylation for the maintenance of non-CG methylation at different loci remain unclear.

Communication of silent information

Epigenetically silent expression states can show remarkable stability throughout mitosis and meiosis, although they can retain the ability to revert to an active state[2]. This gives rise to the concept of the epigenetic allele (epiallele), which is defined as an allele that shows a heritable difference in expression as a consequence of epigenetic modifications and not changes in DNA sequence. For example, hypermethylated (silent) epialleles of *SUPERMAN* (which is involved in floral development) known as *clark kent* are stable during many generations of inbreeding, but they can revert to an unmethylated (active) state at a frequency of ~3% per generation[43]. Another notable characteristic of certain epialleles is their ability to influence other homologous sequences both *in cis* and

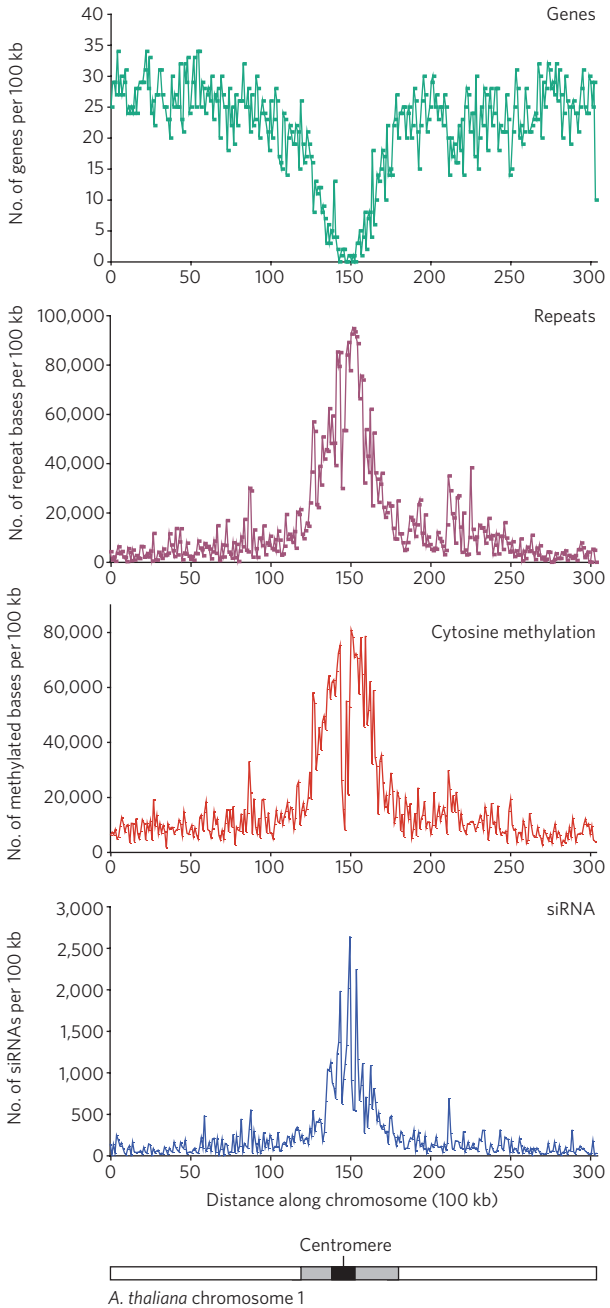


Figure 1 | The epigenetic 'landscape' of *A. thaliana*. The relative abundance of genes (number of annotated genes[11]), repeats (repeat bases per 100 kb; ref. 11), cytosine methylation (methylated bases per 100 kb; ref. 11) and siRNAs (cloned siRNAs per 100 kb; ref. 20) is shown for the length of *A. thaliana* chromosome 1, which is ~30 Mb. Numbers on the *x* axis represent 100-kb windows along the chromosome. A diagram of chromosome 1 is also shown, with white bars indicating euchromatic arms, grey bars indicating pericentromeric heterochromatin and the black bar indicating the centromeric core. (Figure courtesy of X. Zhang, University of California, Los Angeles.)

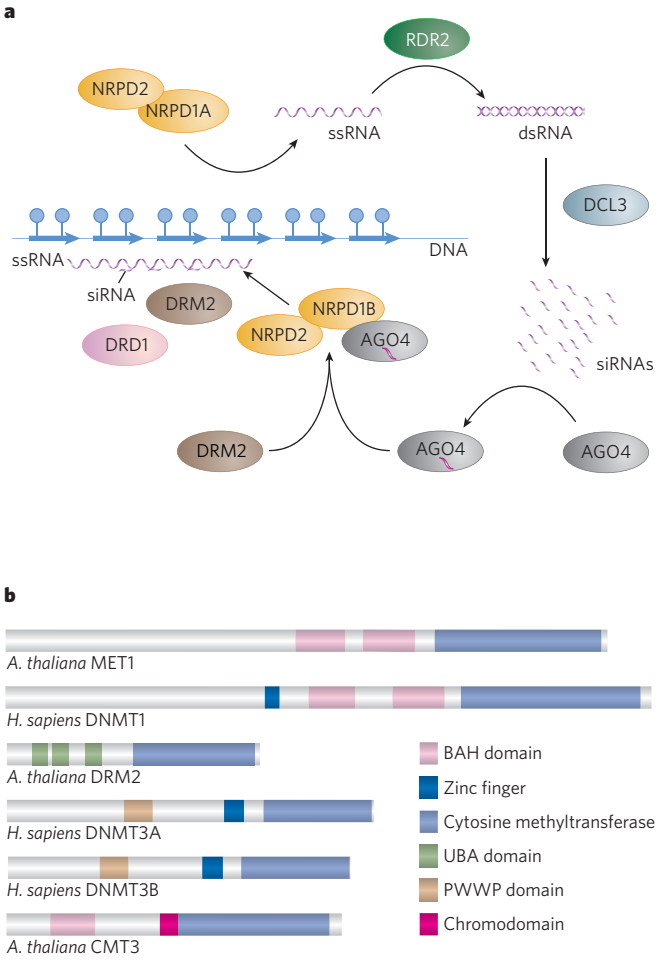


Figure 2 | RNA-directed DNA methylation. a, Putative pathway for RNA-directed DNA methylation in *A. thaliana*. Target loci (in this case tandemly repeated sequences; coloured arrows) recruit an RNA polymerase IV complex consisting of NRPD1A and NRPD2 through an unknown mechanism, and this results in the generation of a single-stranded RNA (ssRNA) species. This ssRNA is converted to double-stranded RNA (dsRNA) by the RNA-dependent RNA polymerase RDR2. The dsRNA is then processed into 24-nucleotide siRNAs by DCL3. The siRNAs are subsequently loaded into the PAZ- and PIWI-domain-containing protein AGO4, which associates with another form of the RNA polymerase IV complex, NRPD1B–NRPD2. AGO4 that is 'programmed' with siRNAs can then locate homologous genomic sequences and guide the protein DRM2, which has *de novo* cytosine methyltransferase activity. Targeting of DRM2 to DNA sequences also involves the SWI–SNF-family chromatin-remodelling protein DRD1. The NRPD1B–NRPD2 complex might generate a target transcript (ssRNA) to which the AGO4-associated siRNAs can hybridize. Given that siRNAs homologous to some loci are absent in *drm2* mutants and *ago4* mutants, it is possible that DNA methylation (blue circles) also stimulates siRNA generation and reinforces silencing. **b**, DNA methyltransferase structure and function. Plant and mammalian genomes encode homologous cytosine methyltransferases, of which there are three classes in plants and two in mammals. *A. thaliana* MET1 and *Homo sapiens* (human) DNMT1 both function to maintain CG methylation after DNA replication, through a preference for hemimethylated substrates, and both have amino-terminal bromo-adjacent homology (BAH) domains of unknown function. *De novo* DNA methylation is carried out by the homologous proteins DRM2 (in *A. thaliana*) and DNMT3A and DNMT3B (both in *H. sapiens*). Despite their homology, these proteins have distinct N-terminal domains, and the catalytic motifs present in the cytosine methyltransferase domain are ordered differently in DRM2 and the DNMT3 proteins. Plants also have another class of methyltransferase, which is not found in mammals. CMT3 functions together with DRM2 to maintain non-CG methylation. PWWP, Pro-Trp-Trp-Pro motif; UBA, ubiquitin associated.

in trans[2]. One example is paramutation, which was discovered in plants and is defined as allelic interactions that cause a meiotically heritable change in the expression of one of the alleles[8]. *Trans*-phenomena similar to paramutation have also been described in mammals, including at a chimaeric version of the mouse *Rasgrf1* (Ras protein-specific guanine-nucleotide-releasing factor 1) locus that contained the imprinting control region from the insulin-like growth factor 2 receptor gene[44].

One of the best-studied paramutation systems is the maize (*Zea mays*) locus *b1*, which encodes a transcription factor that is required for accumulation of the pigment anthocyanin[8]. The paramutagenic epiallele *B′*, which causes light pigmentation, arises spontaneously at a low frequency from its paramutable parent allele *B-I*, which causes dark pigmentation[45]. *B′* epialleles convert *B-I* alleles to *B′* epialleles when heterozygous with 100% penetrance, and the newly created paramutated *B′* epialleles can pass on their silent state in subsequent crosses[45] (Fig. 3). *B′* epialleles are transcribed at one-twentieth to one-tenth the rate of *B-I* alleles but have identical gene sequences[45,46]. Fine-structure recombination mapping of alleles resulting from a cross between individuals with paramutagenic alleles and those with neutral alleles (which cannot participate in paramutation) enabled the sequences required for paramutation to be defined; these sequences are present as an array of 7 tandem 853-base repeats, which is located ~100 kilobases (kb) upstream of *b1* (refs 45, 46). The sequences are present as a single copy in neutral alleles. Recombinant alleles with three repeats show partial paramutational ability, whereas alleles with seven repeats are fully active in paramutation[45,46]. These repeats were also shown to have a closed chromatin structure and more cytosine methylation in *B′* epialleles than in *B-I* alleles[46]. However, for *B′*, cytosine methylation was found to be established after the silent state, so it is unlikely to be the cause[46]. There are several models of *trans*-communication between alleles, including physical pairing of alleles and transmission of an RNA signal. A model for paramutagenic interactions being mediated by siRNA is supported by the finding that a genetic suppressor of paramutation, *mediator of paramutation1* (*mop1*), encodes the maize orthologue of the RNA-dependent RNA polymerase RDR2 (refs 47, 48). So far, siRNAs homologous to the tandem repeats upstream of *B′* have not been reported, although such repeats are commonly associated with small RNAs[20,49]. The *mop1* gene is also required for silencing transgenes and *Mutator*-like transposons, indicating that RNA-dependent RNA polymerases and siRNAs have a role in heterochromatic silencing in monocotyledonous plants[50]. The detailed relationships between siRNAs, chromatin structure at the repeats upstream of *B′*, and the ability to transfer epigenetic states will be intriguing to determine.

The *A. thaliana* gene *FWA* has similarities to maize *b1* in that it has tandem repeats upstream that, when methylated, cause heritable silencing of expression[51]. Stably hypomethylated *fwa-1* epialleles have been found to be generated spontaneously and in *met1* mutant backgrounds[39,40,51], causing overexpression of the transcription factor FWA and a dominant late-flowering phenotype[51]. In contrast to *B′* epialleles, methylated and unmethylated *fwa* epialleles are not influenced by the presence of one another in heterozygotes[23,49,51]. However, introduction of unmethylated transgenic copies of *FWA* by *Agrobacterium tumefaciens*-mediated transformation leads to efficient *de novo* silencing of the incoming transgene, in a process that depends on both DRM2 and the RNA-directed DNA-methylation RNAi pathway[22,23] (Fig. 3). Intriguingly, an unmethylated *FWA* transgene obtained after transformation into a *drm2* mutant does not become remethylated after outcrossing to wild-type *A. thaliana*[22,23]. This finding suggests that, during the transformation process, there is a 'surveillance' window when the incoming *FWA* transgene is competent to be silenced. *A. tumefaciens* targets the female gametophyte (which is haploid) during transformation, but introduction of *FWA* into *DRM2/drm2* heterozygotes revealed that the silencing window must be present after fertilization[49]. Structure–function analysis of an *FWA* transgene showed that the upstream tandem repeats are necessary and sufficient for transformation-dependent silencing and were also found to produce homologous siRNA[49]. Interestingly, the efficiency by which an incoming

FWA transgene is silenced can be influenced by the methylation state of endogenous *FWA*[49]. Whereas introduction of an *FWA* transgene into a background in which the endogenous *FWA* gene is methylated leads to extremely efficient silencing of the transgene, transformation into the *fwa-1* background, which contains an unmethylated endogenous gene, leads to inefficient methylation and silencing of the *FWA* transgene[49] (Fig. 3). Furthermore, an introduced transgene can occasionally cause silencing of the unmethylated *fwa-1* endogenous gene[49]. These results reveal extensive communication between the transgenic and endogenous *FWA* gene copies during transformation, and this communication depends on the DNA methylation state of the endogenous gene. Surprisingly, these differences between *fwa-1* epialleles are not accounted for by siRNA production, because the repeat-derived siRNAs accumulate equally in plants with wild-type *FWA* and those with *fwa-1* (ref. 49). Hence, recruitment of siRNA machinery to a locus is not always sufficient for RNA-directed DNA methylation and probably also requires modifications of chromatin.

Maintenance of silencing at *FWA* depends mainly on CG methylation, because *met1* alleles generate hypomethylated *fwa-1* epialleles at a high frequency[39,40]. Although the tandem repeats upstream of *FWA* are also methylated at non-CG sequences, loss of this methylation in *drm1 drm2 cmt3* triple mutants does not cause reactivation and late flowering[37]. Genome-wide analysis of cytosine methylation and transcription in *drm1 drm2 cmt3* triple mutants has identified genes with methylated promoters, the expression of which depends strongly on DRM- and CMT3-mediated non-CG methylation[11]. These methylated genes might be responsible for the developmental phenotypes of *drm1 drm2 cmt3* triple mutants, which include misshapen leaves and reduced stature[27,37].

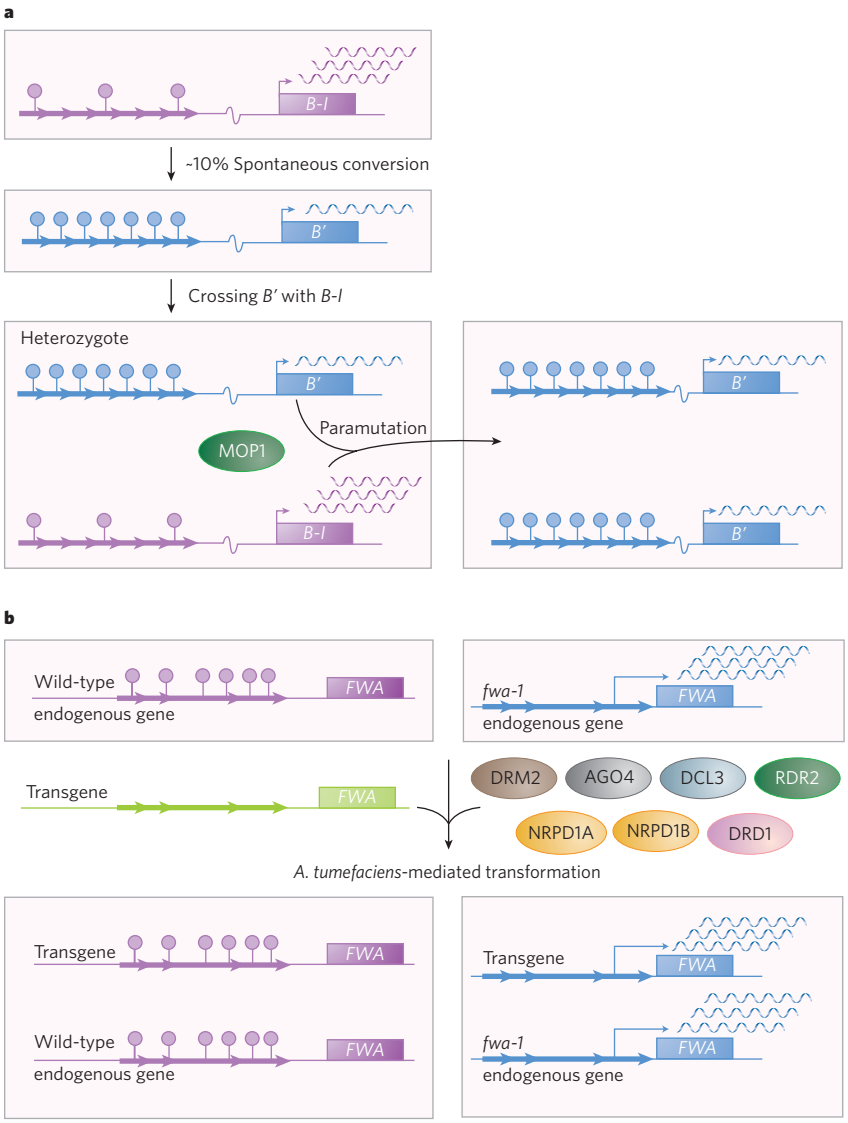


Figure 3 | *Trans*-epiallele interactions at *b1* and *FWA*. **a**, Paramutation at the *b1* locus in maize. The *B-I* allele (pink) of the *b1* gene in maize has an upstream tandem-repeat region (coloured arrows) and spontaneously gives rise to silenced *B′* epialleles (blue) at a low frequency. *B′* epialleles are more heavily methylated at cytosine bases in the repeat region and are less frequently transcribed. When the *B′* epiallele is brought together with a new copy of *B-I* by crossing of maize plants, the *B-I* allele is paramutated to a silenced *B′* state with 100% penetrance. *Trans*-communication between epialleles requires MOP1, the maize homologue of *A. thaliana* RDR2, suggesting that siRNA-mediated silencing might be involved in the conversion of *B-I* to *B′*. **b**, *De novo* silencing of *FWA* transgenes in wild-type and *fwa-1 A. thaliana*. The *FWA* gene in wild-type *A. thaliana* (pink) is methylated at cytosine bases in a pair of tandem repeats in its promoter, silencing its expression. Mutations that decrease DNA methylation give rise to hypomethylated *fwa-1* epialleles (blue), which overexpress the transcription factor FWA, thereby causing late flowering. Introduction of an unmethylated *FWA* transgene (green) by *A. tumefaciens*-mediated transformation of wild-type plants results in efficient methylation and silencing of the incoming transgene. This process depends on DRM2, AGO4, DCL3, RDR2, NRPD1A, NRPD1B and DRD1. By contrast, transformation of an *fwa-1* background results in inefficient silencing of the transgene, indicating that the methylation state of endogenous *FWA* is important for transgene silencing.

In contrast to the independently segregating epialleles that arise in *met1* mutants (as a result of the stable loss of CG methylation)[39,40,51], backcrossing *drm1 drm2 cmt3* triple mutants to wild-type plants or reintroducing either *DRM2* or *CMT3* by transformation immediately rescues these morphological phenotypes[27]. This finding suggests that non-CG methylation can be more easily re-established, possibly allowing flexible regulation of genes. However, it is unclear how commonly this type of regulation is used, because few examples of DNA-methylation-regulated plant genes have been described.

Silencing through time and development

The life cycles of plants differ from those of animals in that the products of meiosis undergo mitotic proliferation to form multicellular gametophytes (that is, the embryo sac and the pollen in flowering plants). The embryo sac (female) contains an egg cell, which is haploid, and this is fertilized by a sperm nucleus, which is also haploid, to form a diploid embryo. A second sperm nucleus fertilizes the central cell, which is diploid, to form triploid endosperm, an extra-embryonic tissue that has a supportive role during embryogenesis. The central cell and the endosperm show parent-of-origin-dependent monoallelic expression, or imprinting, which is important for proper seed development[52]. For example, in *A. thaliana*, the tandem repeats of maternal *FWA* alleles are specifically demethylated in the central cell and the endosperm, leading to expression of *FWA* in these tissues[53]. Demethylation and activation of *FWA* depend on maternal expression of the gene encoding the DNA glycosylase–lyase DEMETER (DME), which can directly excise the base 5-methylcytosine[54–56]. Because the endosperm is a terminally differentiating extra-embryonic tissue, this mechanism does not necessitate remethylation of *FWA*[53]. This is in contrast to mammals, in which demethylation of imprinted genes occurs in primordial germ cells (the cells that ultimately generate the germ line) and is followed by germline-specific remethylation and silencing (see page 425). Other imprinted genes such as *MEA* and *FERTILIZATION-INDEPENDENT SEED 2* also have cytosine-methylated regions in their promoters that are associated with maternally restricted expression[55,57]. However, only for *FWA* has it been shown that differential methylation of particular sequences is required for the regulation of imprinting[53,58].

Cytosine demethylation is also likely to have an important role in the control of silencing in situations other than gametophytic generation and imprinting. *DME* belongs to a small *A. thaliana* gene family that includes the somatically expressed gene *REPRESSOR OF SILENCING 1* (*ROS1*)[54,59]. Mutations in *ROS1* have been shown to increase RNA-directed DNA methylation, and ROS1 has been shown to function as a cytosine demethylase[56,59,60]. Together, these exciting discoveries have defined a long-sought cytosine demethylation pathway, and they raise many interesting questions. For example, to what extent are genomic methylation patterns balanced by the targeting of *de novo* DNA methyltransferases and DNA glycosylases? Furthermore, there are indications of a similar mechanism for cytosine demethylation in vertebrates[61,62].

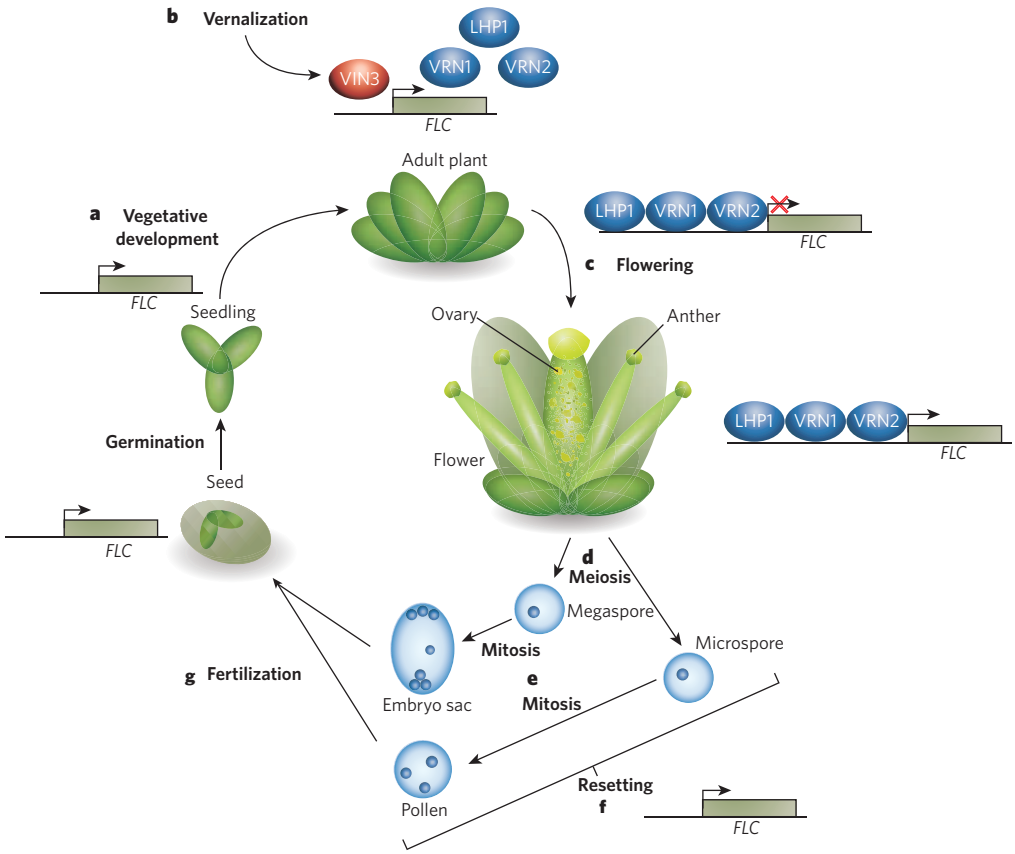


Figure 4 | PcG-protein-mediated silencing throughout the *A. thaliana* life cycle. The activation state of the PcG protein target *FLC* is illustrated throughout the plant life cycle. **a**, *FLC* is transcriptionally active in seeds and seedlings, preventing the plant from flowering and prolonging vegetative development. **b**, Exposure to a long period of cold (that is, vernalization) results in the expression of VIN3 (red), which initiates repression of *FLC* transcription, and the binding of the PcG protein VRN2, as well as VRN1 and LHP1 (blue). In this process, chromatin at *FLC* is epigenetically modified by the trimethylation of H3K27. **c**, After warmer temperatures return, *FLC* repression is maintained, allowing flowering to be induced by other cues. **d**, During flower development, the anthers and ovaries are sites of meiotic differentiation, giving rise to haploid cells known as microspores and megaspores, respectively. **e**, These meiotic products undergo mitotic proliferation to form the multicellular embryo sac and pollen gametophytes. **f**, PcG-protein-mediated repression at *FLC* is removed during an undefined resetting process. **g**, Then, the pollen contributes sperm nuclei to the embryo sac, and these fertilize the haploid egg cell and diploid central cell (not shown), forming the embryo and endosperm (respectively) in a new seed, in which *FLC* is re-expressed.

Other examples of imprinted genes are maize *fertilization-independent endosperm1* (*fie1*) and *fie2*, which show monoallelic expression from the maternal allele during endosperm development. This is reflected by the promoters of the silent paternal alleles having differentially methylated regions (DMRs)[63,64]. Analysis of DMR methylation of *fie* alleles in sperm, egg and central cells showed interesting differences in the mechanism for imprinting *fie1* and *fie2* (ref. 64). The DMR of *fie1* is heavily methylated in all three cell types, but the maternal alleles in the central cell (which contribute to the endosperm) become specifically demethylated, resembling the imprinting mechanism described for *A. thaliana FWA*[64]. By contrast, the DMR of *fie2* is unmethylated in all gametes, although the paternal allele becomes methylated *de novo* in the endosperm. Furthermore, the *fie2* DMR also showed extensive non-CG methylation, which is consistent with a DRM2-type-mediated RNA-directed DNA methylation process[64]. A further instance of potential gene regulation by *de novo* DNA methylation is provided by the *Brassica rapa SP11* locus, which encodes a pollen self-incompatibility determinant[65]. The *B. rapa* self-incompatibility phenotype is controlled by dominance relationships between *S*-haplotypes, and recessive *SP11* alleles were found to be specifically methylated *de novo* and silenced in the anther tapetal tissues[65]. It will be interesting to determine the prevalence of such instances of tissue-specific gene regulation by DNA methylation.

In addition to the gametophytic tissues being an important location for the establishment of imprinted gene expression, they also maintain pre-existing patterns of cytosine methylation. Evidence that silencing is important during gametophytic generation is provided by null *met1* alleles in *A. thaliana*, which produce hypomethylated epialleles even when the individual is heterozygous for the null allele[40]. This is caused by loss of cytosine methylation in the gametophytes of *met1* mutants, a loss that is greater when *met1* is inherited through the female gametophyte than the male[40]. This difference is probably accounted for by the female gametophyte (that is, the embryo sac) undergoing one more postmeiotic round of DNA replication before fertilization than the male gametophyte (that is, the pollen)[40].

A different epigenetic system used to developmentally silence genes during plant life cycles involves Polycomb group (PcG) proteins[66]. A conserved complex known as Polycomb repressive complex 2 (PRC2) functions to maintain patterns of gene repression in both plants and animals, using H3K27 methylation[66] (see page 425). However, in plants, there are several PRC2 complexes, with overlapping subunit compositions, specialized for distinct developmental roles[66]. For example, the PcG proteins have an important role in the regulation of imprinted gene expression. *A. thaliana MEA*, which is a homologue of *Drosophila melanogaster Enhancer of zeste*, shows maternally imprinted expression[67]. An important component of *MEA* imprinting is repression of the paternal *MEA* allele in the endosperm, and this process has been found to involve *MEA* autoregulation, using H3K27 trimethylation[55,68,69]. Interestingly, the mammalian PcG protein EED (embryonic ectoderm development) has also been shown to have an important role in the control of imprinted gene expression[70].

Another well-understood example of PcG-protein-mediated regulation in plants involves silencing of the floral-repressor gene *FLOWERING LOCUS C* (*FLC*) during the vernalization response in *A. thaliana*[71–73] (Fig. 4). Expression of *FLC*, which encodes a MADS-box-containing transcription factor, delays flowering and can be silenced by exposure of the plant to long periods of cold (that is, vernalization)[71–73]. In nature, this cold treatment occurs in winter and leads to flowering in favourable spring conditions. After the cold signal has been removed, *FLC* silencing is stable[71–73]. Mutations in the *VERNALIZATION 2* (*VRN2*) gene, which encodes a homologue of the *D. melanogaster* PcG protein Suppressor of zeste 12, cause late flowering after vernalization as a result of high levels of *FLC* expression[72]. Interestingly, *vrn2* mutants can silence *FLC* expression during the cold but fail to maintain this repression after the cold signal has been removed[72]. VRN2 is also required for acquisition of H3K27 dimethylation and trimethylation at *FLC* during vernalization, consistent with the known functions of PRC2 in maintaining patterns of gene repression[71,73,74].

The mechanism by which the vernalization-specific PcG-protein complex is recruited to *FLC* is not well understood but is known to require the PHD-finger-domain-containing protein VERNALIZATION INSENSITIVE 3 (VIN3)[73]. Because *VIN3* expression is induced after cold treatment, this protein might be a component of the signalling pathway that recruits PcG-protein-mediated repression to *FLC*[73] (Fig. 4). Recently, the *A. thaliana* homologue of *D. melanogaster* Heterochromatin protein 1 (HP1) — LIKE HETEROCHROMATIN PROTEIN 1 (LHP1; also known as TFL2) — was found to be required for the maintenance of *FLC* silencing after vernalization[75,76]. LHP1 becomes associated with the silenced *FLC* locus, a process that depends on an intronic sequence element[76]. The role of LHP1 in the repression of PcG-protein-regulated genes differs markedly from the main role of animal HP1 in heterochromatic silencing (see page 399). The DNA-binding protein VRN1 is also required for the maintenance of *FLC* silencing and associates with mitotic chromosomes[75,77]. Interestingly, VRN1 is absent from meiotic chromosomes of developing pollen[75]. One speculation is that this absence is associated with the resetting of *FLC* expression, which leads to a requirement for vernalization, at the start of each generation. Indeed, all PcG-protein-mediated silencing might be reset at some point during meiosis or gametogenesis, through an unknown mechanism (Fig. 4).

Conclusions

Plants continue to be excellent systems for the study of epigenetics, and their silencing mechanisms have marked similarities to those of mammals. An advantage of using plants is that they are tolerant of genome stresses, such as large losses of DNA methylation and changes in chromosome number. The elegant genetic tools available for organisms such as maize and *A. thaliana* are facilitating the dissection of epigenetic control. Recent advances such as the development of whole-genome microarrays and high-throughput sequencing are allowing the generation of large-scale data sets for epigenetic modifications and small RNAs that are extending our view to a genome-wide scale. Together, these approaches should enable major advances in our understanding of epigenetics to be made using plant systems: for example, how specific chromatin modifications are established and maintained, how they influence one another, and the extent to which they are used throughout the genome. This work should provide important insight for fields as diverse as cancer biology, development and evolution. ■

1. Gregory, T. R. The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann. Bot. (Lond.)* **95,** 133–146 (2005).
2. Hall, I. M. & Grewal, S. I. in *RNAi: A Guide to Gene Silencing* (ed. Hannon, G. J.) 205–232 (Cold Spring Harbor Laboratory Press, Woodbury, 2003).
3. Bernstein, B. E., Meissner, A. & Lander, E. S. The epigenome. *Cell* **128,** 669–681 (2006).
4. Bernard, P. *et al.* Requirement of heterochromatin for cohesion at centromeres. *Science* **294,** 2539–2542 (2001).
5. Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441,** 87–90 (2006).
6. Liu, J., He, Y., Amasino, R. & Chen, X. siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev.* **18,** 2873–2878 (2004).
7. Comfort, N. C. From controlling elements to transposons: Barbara McClintock and the Nobel Prize. *Trends Biochem. Sci.* **26,** 454–457 (2001).
8. Chandler, V. L. & Stam, M. Chromatin conversations: mechanisms and implications of paramutation. *Nature Rev. Genet.* **5,** 532–544 (2004).
9. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286,** 950–952 (1999).
10. Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed *de novo* methylation of genomic sequences in plants. *Cell* **76,** 567–576 (1994).
11. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126,** 1189–1201 (2006).
12. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39,** 61–69 (2007).
13. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815 (2000).
14. Fransz, P. F. *et al.* High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibres. *Plant J.* **9,** 421–430 (1996).
15. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430,** 471–476 (2004).
16. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297,** 1833–1837 (2002).
17. Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, A. J. & Matzke, M. RNA-directed DNA methylation in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99** (suppl. 4), 16499–16506 (2002).

18. Mochizuki, K., Fine, N. A., Fujisawa, T. & Gorovsky, M. A. Analysis of a *piwi*-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell* **110,** 689–699 (2002).
19. Matzke, M., Matzke, A. J. & Kooter, J. M. RNA: guiding gene silencing. *Science* **293,** 1080–1083 (2001).
20. Lu, C. *et al.* Elucidation of the small RNA component of the transcriptome. *Science* **309,** 1567–1569 (2005).
21. Cao, X. *et al.* Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* **13,** 2212–2217 (2003).
22. Cao, X. & Jacobsen, S. E. Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr. Biol.* **12,** 1138–1144 (2002).
23. Chan, S. W. *et al.* RNA silencing genes control *de novo* DNA methylation. *Science* **303,** 1336 (2004).
24. Zilberman, D. *et al.* Role of *Arabidopsis ARGONAUTE4* in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* **14,** 1214–1220 (2004).
25. Henderson, I. R. *et al.* Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nature Genet.* **38,** 721–725 (2006).
26. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2,** e104 (2004).
27. Chan, S. W. *et al.* RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in *Arabidopsis*. *PLoS Genet.* **2,** e83 (2006).
28. Li, C. F. *et al.* An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126,** 93–106 (2006).
29. Pontes, O. *et al.* The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126,** 79–92 (2006).
30. Qi, Y. *et al.* Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443,** 1008–1012 (2006).
31. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299,** 716–719 (2003).
32. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308,** 118–120 (2005).
33. Kanno, T. *et al.* Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nature Genet.* **37,** 761–765 (2005).
34. Onodera, Y. *et al.* Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120,** 613–622 (2005).
35. Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19,** 2030–2040 (2005).
36. Kanno, T. *et al.* Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol.* **14,** 801–805 (2004).
37. Cao, X. & Jacobsen, S. E. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc. Natl Acad. Sci. USA* **99** (suppl. 4), 16491–16498 (2002).
38. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74,** 481–514 (2005).
39. Kankel, M. W. *et al. Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* **163,** 1109–1122 (2003).
40. Saze, H., Mittelsten Scheid, O. & Paszkowski, J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nature Genet.* **34,** 65–69 (2003).
41. Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416,** 556–560 (2002).
42. Malagnac, F., Bartee, L. & Bender, J. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* **21,** 6842–6852 (2002).
43. Jacobsen, S. E. & Meyerowitz, E. M. Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* **277,** 1100–1103 (1997).
44. Herman, H. *et al. Trans* allele methylation and paramutation-like effects in mice. *Nature Genet.* **34,** 199–202 (2003).
45. Stam, M. *et al.* The regulatory regions required for *B'* paramutation and expression are located far upstream of the maize *b1* transcribed sequences. *Genetics* **162,** 917–930 (2002).
46. Stam, M., Belele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* **16,** 1906–1918 (2002).
47. Alleman, M. *et al.* An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442,** 295–298 (2006).
48. Woodhouse, M. R., Freeling, M. & Lisch, D. Initiation, establishment, and maintenance of heritable *MuDR* transposon silencing in maize are mediated by distinct factors. *PLoS Biol.* **4,** e339 (2006).
49. Chan, S. W.-L., Zhang, X., Bernatavichute, Y. V. & Jacobsen, S. E. Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol.* **4,** e363 (2006).
50. Lisch, D., Carey, C. C., Dorweiler, J. E. & Chandler, V. L. A mutation that prevents paramutation in maize also reverses *Mutator* transposon methylation and silencing. *Proc. Natl Acad. Sci. USA* **99,** 6130–6135 (2002).
51. Soppe, W. J. *et al.* The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6,** 791–802 (2000).
52. Gehring, M., Choi, Y. & Fischer, R. L. Imprinting and seed development. *Plant Cell* **16,** S203–S213 (2004).
53. Kinoshita, T. *et al.* One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303,** 521–523 (2004).
54. Choi, Y. *et al.* DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* **110,** 33–42 (2002).
55. Gehring, M. *et al.* DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124,** 495–506 (2006).
56. Morales-Ruiz, T. *et al.* DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Natl Acad. Sci.USA* **103,** 6853–6858 (2006).
57. Jullien, P. E., Kinoshita, T., Ohad, N. & Berger, F. Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. *Plant Cell* **18,** 1360–1372 (2006).
58. Kinoshita, Y. *et al.* Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* **49,** 38–45 (2007).
59. Gong, Z. *et al.* ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell* **111,** 803–814 (2002).
60. Agius, F., Kapoor, A. & Zhu, J. K. Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl Acad. Sci. USA* **103,** 11796–11801 (2006).
61. Barreto, G. *et al.* Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445,** 671–675 (2007).
62. Jost, J. P., Siegmann, M., Sun, L. & Leung, R. Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. *J. Biol. Chem.* **270,** 9734–9739 (1995).
63. Danilevskaya, O. N. *et al.* Duplicated *fie* genes in maize: expression pattern and imprinting suggest distinct functions. *Plant Cell* **15,** 425–438 (2003).
64. Gutierrez-Marcos, J. F. *et al.* Epigenetic asymmetry of imprinted genes in plant gametes. *Nature Genet.* **38,** 876–878 (2006).
65. Shiba, H. *et al.* Dominance relationships between self-incompatibility alleles controlled by DNA methylation. *Nature Genet.* **38,** 297–299 (2006).
66. Kohler, C. & Grossniklaus, U. Epigenetic inheritance of expression states in plant development: the role of Polycomb group proteins. *Curr. Opin. Cell Biol.* **14,** 773–779 (2002).
67. Kinoshita, T., Yadegari, R., Harada, J. J., Goldberg, R. B. & Fischer, R. L. Imprinting of the *MEDEA* polycomb gene in the *Arabidopsis* endosperm. *Plant Cell* **11,** 1945–1952 (1999).
68. Baroux, C., Gagliardini, V., Page, D. R. & Grossniklaus, U. Dynamic regulatory interactions of Polycomb group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis*. *Genes Dev.* **20,** 1081–1086 (2006).
69. Jullien, P. E., Katz, A., Oliva, M., Ohad, N. & Berger, F. Polycomb group complexes self-regulate imprinting of the Polycomb group gene *MEDEA* in *Arabidopsis*. *Curr. Biol.* **16,** 486–492 (2006).
70. Mager, J., Montgomery, N. D., de Villena, F. P. & Magnuson, T. Genome imprinting regulated by the mouse Polycomb group protein Eed. *Nature Genet.* **33,** 502–507 (2003).
71. Bastow, R. *et al.* Vernalization requires epigenetic silencing of *FLC* by histone methylation. *Nature* **427,** 164–167 (2004).
72. Gendall, A. R., Levy, Y. Y., Wilson, A. & Dean, C. The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* **107,** 525–535 (2001).
73. Sung, S. & Amasino, R. M. Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature* **427,** 159–164 (2004).
74. Sung, S., Schmitz, R. J. & Amasino, R. M. A PHD finger protein involved in both the vernalization and photoperiod pathways in *Arabidopsis*. *Genes Dev.* **20,** 3244–3248 (2006).
75. Mylne, J. S. *et al.* LHP1, the *Arabidopsis* homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of *FLC*. *Proc. Natl Acad. Sci. USA* **103,** 5012–5017 (2006).
76. Sung, S. *et al.* Epigenetic maintenance of the vernalized state in *Arabidopsis thaliana* requires LIKE HETEROCHROMATIN PROTEIN 1. *Nature Genet.* **38,** 706–710 (2006).
77. Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R. & Dean, C. Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. *Science* **297,** 243–246 (2002).

Acknowledgements We thank S. Chan, C. Fei Li, K. Niakan, M. Ong and all members of the Jacobsen laboratory for useful comments and discussion. We apologize to colleagues whose research we did not have space to discuss. I.R.H. was supported by a long-term fellowship from the European Molecular Biology Organization, a Special Fellow grant from The Leukemia & Lymphoma Society, and a grant from the National Institutes of Health. S.E.J is an investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to S.E.J. (jacobsen@ucla.edu).

Stability and flexibility of epigenetic gene regulation in mammalian development

Wolf Reik[1]

During development, cells start in a pluripotent state, from which they can differentiate into many cell types, and progressively develop a narrower potential. Their gene-expression programmes become more defined, restricted and, potentially, 'locked in'. Pluripotent stem cells express genes that encode a set of core transcription factors, while genes that are required later in development are repressed by histone marks, which confer short-term, and therefore flexible, epigenetic silencing. By contrast, the methylation of DNA confers long-term epigenetic silencing of particular sequences — transposons, imprinted genes and pluripotency-associated genes — in somatic cells. Long-term silencing can be reprogrammed by demethylation of DNA, and this process might involve DNA repair. It is not known whether any of the epigenetic marks has a primary role in determining cell and lineage commitment during development.

Development is, by definition, epigenetic. Differences in the programmes of gene expression that result in the development of different organs and tissues occur without changes to the sequence of our DNA (with one or two exceptions). There is nothing mysterious in this concept; subsets of the ~30,000 genes in our genome are active in different tissues and organs, depending on their regulation by different sets or combinations of transcription factors. This implies that if we were to take all of the transcription factors that activate genes in a liver cell and transfer them to a brain cell (while inactivating all brain-specific transcription factors), then the brain cell would turn into a liver cell.

A recent study provides tantalizing insight into this concept of epigenetic control of development. Takahashi and Yamanaka identified four transcriptional regulators that when expressed in fibroblasts, resulted in these cells being reprogrammed to become embryonic stem (ES)-like cells[1]. Extending this concept a little further, in somatic-cell nuclear transfer, the nucleus of a somatic cell from an adult individual is transplanted into an oocyte from which the nucleus has been removed, resulting in reprogramming of the adult nucleus and therefore successful development of the cloned animal.

Cloning, however, is inefficient, because most (if not all) cloned animals have epigenetic defects, particularly in DNA methylation. Therefore, our lack of understanding of how epigenetic marks are reprogrammed is a key obstacle to cloning[2]. Similarly, the reprogramming of fibroblasts to become ES-like cells is a rare event *in vitro*, and epigenetic defects such as lack of demethylation of the *Oct4* (also known as *Pou5f1*) promoter, affecting expression of the encoded transcription factor, have been noted in these ES-like cells[1].

These observations highlight that, in addition to transcription factors, changes in gene expression during development are accompanied or caused by epigenetic modifications[2–7], such as methylation of DNA at CpG sequences (in vertebrates[4,5]), modification of histone tails[6] and the presence of non-nucleosomal chromatin-associated proteins[7]. Therefore, as development and differentiation proceed, differentiated cells accumulate epigenetic marks that differ from those of pluripotent cells, and differentiated cells of different lineages also accumulate different marks.

In this review, I focus on the role of epigenetic regulation in development, particularly comparing the short-term flexibility of certain epigenetic marks (which can be removed before a cell divides or within very few cell divisions) with the long-term stability and heritability of other marks (which can be maintained for many divisions) (Fig. 1). During the early stages of development, genes that are required later in development are transiently held in a repressed state by histone modifications, which are highly flexible and easily reversed when expression of these genes is needed. During differentiation, genes that are crucial for pluripotency are silenced by histone modifications, as well as by DNA methylation. Some of these genes are also silent in mature germ cells, meaning that epigenetic marks probably need to be reversed rapidly after fertilization to allow re-expression of pluripotency-associated genes in the next generation. By contrast, long-term silencing of transposons and imprinted genes — which is based on DNA methylation — needs to be stably maintained from the gametes into the early embryo and the adult organism. Methylation of imprinted genes can only be erased in primordial germ cells (PGCs), the cells that ultimately give rise to the germ line. Probably because there is a requirement for both removing epigenetic marks and retaining epigenetic marks between generations, epigenetic information can sometimes be inherited across multiple generations. In this review, I address how the fascinating interplay between transcription factors and epigenetic factors is beginning to provide an explanation for how pluripotency and development are regulated.

Flexibility for developmental gene regulation

In this section, three issues are addressed. First, are differentiation-specific genes held in an epigenetically silenced manner in pluripotent cell types, in order to be activated later? And is the removal of epigenetic marks from these genes needed for their activation? Second, are pluripotency-associated genes epigenetically inactivated in differentiated cell types? This inactivation could, in principle, be irreversible, because somatic cell types are not required to give rise to pluripotent cells. One exception is the germ line, where reactivation of pluripotency-associated genes is needed at the initial stages of development; however, later, the silencing of these genes is essential for the differentiation of mature germ cells. And therefore, third, is the removal of 'permanent' silencing marks from the gametic genomes after fertilization crucial to activate essential genes, such as pluripotency-associated genes, early in development?

[1]Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB22 3AT, UK.

There is recent evidence for the first type of epigenetic regulation: that is, the temporary inactivation of differentiation-specific genes in pluripotent cell types (Fig. 2a). Genes that are required during development and differentiation — for example, those in the homeobox (*Hox*), distal-less homeobox (*Dlx*), paired box (*Pax*) and sine-oculis-related homeobox (*Six*) gene families — are held repressed in pluripotent ES cells by the Polycomb group (PcG)-protein repressive system in mice and humans. This system marks the histones associated with these genes by inducing methylation of the lysine residue at position 27 of histone H3 (H3K27)[8–10]. ES cells that lack EED (embryonic ectoderm development), a component of the PcG-protein repressive complex (PRC), have partly derepressed developmental genes and are prone to spontaneous differentiation[8,10]. Interestingly, some developmental genes are present within 'bivalent' chromatin regions, which contain both inactivating marks (methylated H3K27) and activating marks (methylated H3K4)[9,11]. This could indicate that after the repressive marks have been removed (when expression of the components of PRCs are downregulated during differentiation), these genes are automatically poised for transcriptional activation through the H3K4 methylation mark. It is important to note that epigenetic silencing by PRCs might be mitotically heritable (through an unknown mechanism)[7], but these marks could presumably be rapidly removed by enzymatic demethylation of H3K27 (by an unidentified demethylase)[12]. The H3K27 methylation mark occurs mostly outside the context of DNA methylation. In contrast to the terminal silencing achieved by DNA methylation (discussed later), developmental genes that are silenced by PRCs in pluripotent tissues require repressive marks to be rapidly and flexibly removed when differentiation begins. Strikingly, in cancer cells, the genes targeted by PRCs often become DNA methylated, which might result in a more permanent locking-in of a 'pluripotent' state in cancer stem cells[13].

The second type of epigenetic regulation to be considered is whether pluripotency-associated genes are epigenetically inactivated in differentiated cell types. Several genes that are required for early development or for germ-cell development only — for example, those that encode pluripotency-sustaining transcription factors (such as OCT4 and NANOG) — are known to be expressed by ES cells but silenced on the differentiation of these cells, with a defined kinetics of acquiring repressive histone modifications and DNA methylation[14] (Fig. 2b). Silencing by both histone modifications and DNA methylation in somatic tissues seems to be typical of this group of genes and of those that encode cancer–testis antigens, which are expressed during spermatogenesis[15]. It is probable that this permanent type of epigenetic silencing safeguards against accidental expression of these genes in differentiated cells, because that might lead to dedifferentiation and, perhaps, to a predisposition to

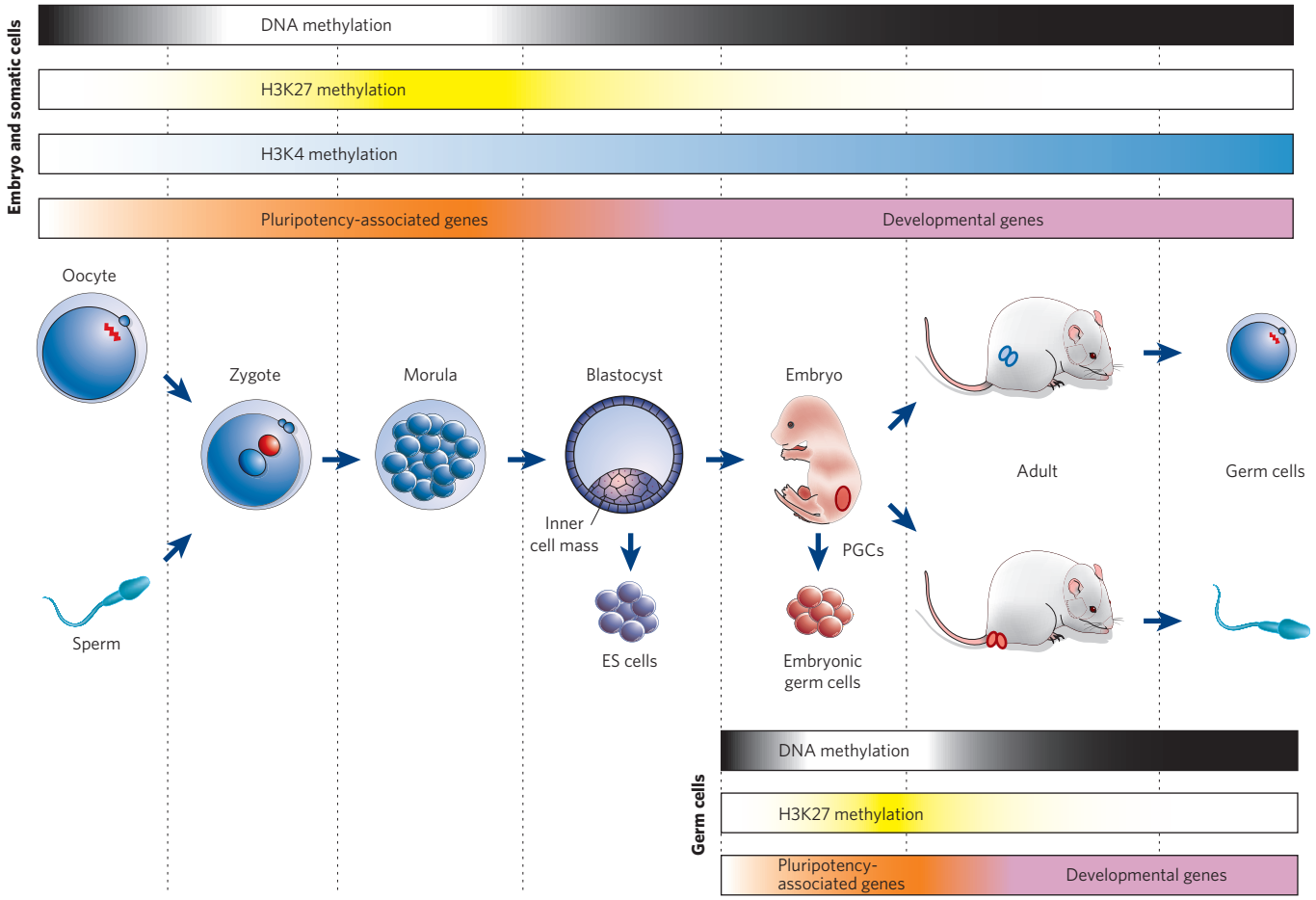


Figure 1 | Epigenetic gene regulation during mammalian development. Key developmental events are shown together with global epigenetic modifications and gene-expression patterns. Very early in development, DNA methylation is erased. In addition, pluripotency-associated genes begin to be expressed, and developmental genes are repressed by the PcG protein system and H3K27 methylation. During the differentiation of pluripotent cells such as ES cells, pluripotency-associated genes are repressed, potentially permanently, as a result of DNA methylation. At the same time, developmental genes begin to be expressed, and there is an increase in H3K4 methylation. During the early development of PGCs, DNA methylation and repressive histone modifications (such as H3K9 methylation) are also erased. Pluripotency-associated genes are re-expressed during a time window that allows embryonic germ cells to be derived in culture. Imprinted genes are demethylated during this period, and developmental genes are expressed afterwards. Flexible histone marks such as H3K27 methylation enable developmental genes to be silenced for a short time in pluripotent cells. By contrast, DNA methylation enables the stable silencing of imprinted genes, transposons and some pluripotency-associated genes.

cancer[16]. Consequently, these genes are difficult to reactivate in cloned embryos because of inefficient reprogramming of repressive marks, particularly of DNA methylation[17].

Special epigenetic regulation needs to occur in PGCs developing in the early post-implantation embryo[18]. Because these cells emerge from cell types in the egg cylinder that are already on the way to lineage commitment and differentiation, the somatic gene-expression programme needs to be suppressed. One of the key regulators of this process is BLIMP1 (B-lymphocyte-induced maturation protein 1), which associates with the arginine methyltransferase PRMT5. PRMT5 might partly repress *Hox*-family genes and other somatic genes in PGCs[19] (Fig. 2c). Pluripotency-associated genes and genes that have later roles in germ-cell development can also be repressed by DNA methylation (Fig. 2b). So genes such as *Mvh* (also known as *Ddx4*), *Dazl* (deleted in azoospermia-like) and *Sycp3* (synaptonemal complex protein 3) are methylated in early PGCs and begin to be expressed after the erasure of DNA methylation[20], which occurs between embryonic day (E) 8.0 and E12.5 in PGCs. Interestingly, pluripotency-associated genes such as *Nanog* also begin to be reactivated at these stages, but it is not known whether this involves demethylation of DNA. PGCs at these stages have similar properties to pluripotent cells, including the ability to form embryonic germ cells in culture[21]. These studies are important because they are the first to show that in some developmental situations, removal of epigenetic marks (H3K27 methylation in the ES-cell study, and DNA methylation in the PGC study) could be crucial for the activation of developmental genes. Whether DNA methylation in PGCs is erased by an active or a passive mechanism is unclear (discussed later). The promoters of the genes that undergo 'developmental' demethylation (for example, *Mvh*, *Dazl* and *Sycp3*) contain CpG islands, as do the differentially methylated regions (DMRs) of imprinted genes, which also undergo demethylation at these stages of PGC development. I am not aware of any reports of demethylation of CpG islands during development other than in PGCs or in the zygote and pre-implantation embryo (discussed later). Methylation of CpG islands might only be removable under exceptional circumstances.

Some key pluripotency-associated genes (such as *Oct4* and *Nanog*) are epigenetically inactivated at later stages of gametogenesis and in the mature gametes, including by DNA methylation. Therefore, after

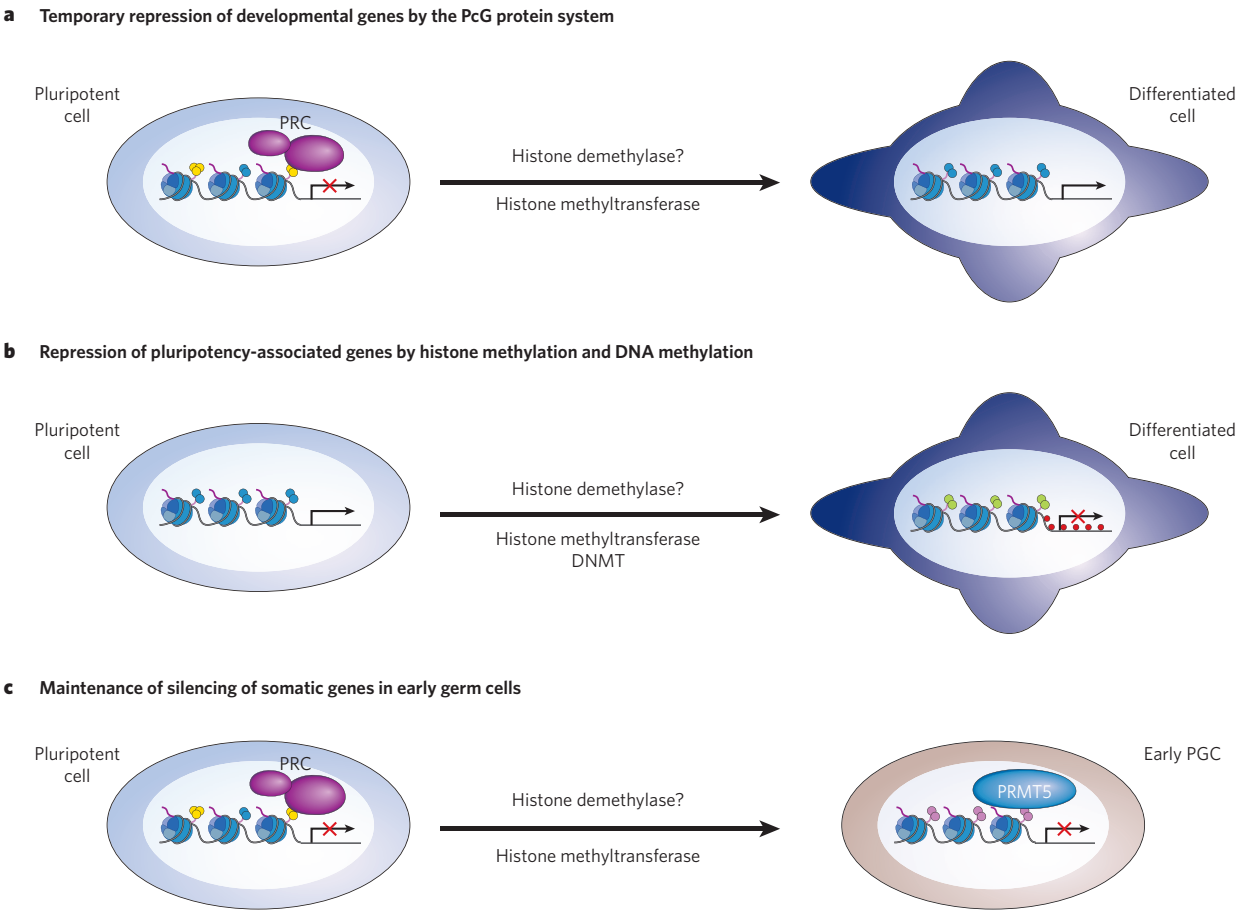


Figure 2 | Epigenetic regulation of pluripotency-associated genes and developmental genes during the differentiation of somatic cells and germ cells. The expression or repression of pluripotency-associated genes and developmental genes is indicated, and the associated modifications of the histone tails and/or DNA are represented by different colours. **a**, In pluripotent cells, the repression of genes that are needed later in development is flexible and can involve the PcG-protein repressive system. Silent developmental genes can be marked by both H3K27 methylation (yellow) and H3K4 methylation (blue), possibly allowing rapid gene activation after loss of repression by PcG-protein-containing repressive complexes (PRCs). Whether the loss of H3K27 methylation involves a histone demethylase is unknown. Further increases in H3K4 methylation might be required for proper developmental gene expression. **b**, Pluripotency-associated genes are stably silenced during differentiation, through histone methylation and DNA methylation. For example, genes such as *Oct4* and *Nanog* are silenced during ES-cell differentiation, and this process can involve both histone methylation (such as methylation of H3K9 mediated by G9A; also known as EHMT2) (green) and DNA methylation (red). Whether a histone demethylase is required for the removal of H3K4 methylation is unknown. **c**, For germ-cell development, the repression of somatic genes needs to be maintained in early germ cells, and this process might involve histone arginine methylation (pink). *Hox*-family genes and other developmental genes remain silent in early germ cells; some of this silencing might require histone arginine methylation brought about by PRMT5.

fertilization, the repressive epigenetic marks might need to be removed for transcriptional activation of these genes and correct early lineage development to take place (discussed later).

Stability for transposon silencing and imprinting

In contrast to developmental genes, which need to be epigenetically regulated with flexibility, transposons (if possible) need to be silenced completely and stably (at least from the perspective of the host) to prevent them from moving around in the genome and potentially causing mutations[22]. Therefore, many transposon families are both methylated themselves and marked by repressive histone modifications (such as H3K9 methylation), and these marks are important for the heritable silencing of transposons. Some transposon families (such as intracisternal A particles; IAPs) are also resistant to the erasure of DNA methylation in the zygote and in PGCs, possibly resulting in epigenetic inheritance across generations (discussed later).

Imprinted genes are a class of mammalian genes with possible mechanistic relationships to transposons[23], in that CpG islands in their promoters become methylated and in that silencing relies on long-term epigenetic stability. In imprinted genes (and transposons), DNA methylation is introduced during either oogenesis or spermatogenesis, by the *de novo* methyltransferase DNA methyltransferase 3A (DNMT3A) and its cofactor DNMT3-like (DNMT3L)[24,25] (Fig. 3a). How particular imprinted genes are selected for *de novo* methylation during oogenesis or spermatogenesis is not understood, although this targeting could involve pre-existing histone marks[26]. After fertilization, the methylation of imprinted-gene DMRs is maintained by DNMT1o (the oocyte form of DNMT1) for one division cycle during very early pre-implantation development[27] and then by DNMT1s (the somatic form of DNMT1) in embryonic and adult tissues[28].

Imprinted genes can be directly silenced by methylation of DMRs (which often contain CpG islands) that overlap the promoter. More frequently, however, imprinted genes occur in clusters, and there is usually a single DMR that is methylated in the germ line and is responsible for regulating gene silencing in the rest of the cluster. So far, there are two distinct models for how, after fertilization, imprinted genes are silenced through the action of nearby unmethylated DMRs. First, the DMR overlaps the promoter of a long, non-coding, unspliced, nuclear RNA[29,30]. The presence of the unmethylated and expressed copy of the non-coding RNA results in the silencing of linked genes, a process that involves repressive histone modifications[31,32]. It is unclear how the presence of the non-coding RNA leads to gene silencing *in cis*. In one model, repressive complexes (for example, PRCs) might be targeted during transcription[33]. Alternatively, the RNA might 'coat' the region to be inactivated, similarly to how *Xist* RNA (inactive X-specific transcripts) coats the inactive X chromosome[31,34]. This might establish a physical structure from which RNA polymerase II (Pol II) is excluded, resulting in transcriptional silencing[35] (Fig. 3b). In one case of silencing mediated by an imprinted non-coding RNA, the developmental kinetics of inactivation are markedly similar to those of imprinted X-chromosome inactivation. Both

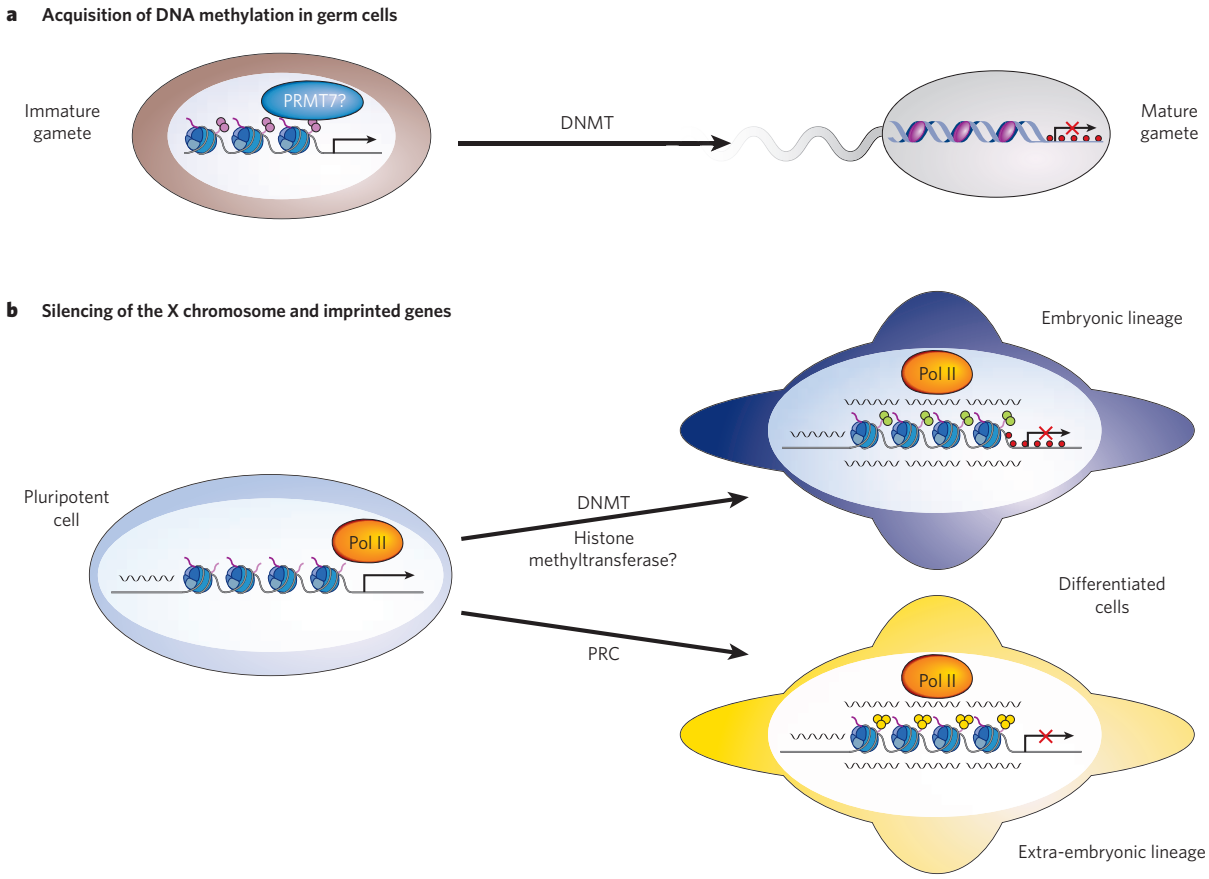


Figure 3 | Developmental regulation of imprinting and X-chromosome inactivation. a, During germ-cell development, selected imprinted genes and transposons become methylated. This process depends on *de novo* methyltransferases such as DNMT3A and its cofactor DNMT3L. It is possible that the targeting of DNA methylation requires arginine methylation of histones, carried out by PRMT7. Mature male germ cells have chromatin that is largely based on non-histone proteins known as protamines (dark pink); this alters the packaging of the DNA. **b**, Expression of non-coding RNAs (wavy black line) *in cis* can result in the silencing of adjacent genes as a consequence of the physical exclusion of Pol II and the acquisition of histone modifications and/or DNA methylation, depending on the embryonic lineage. DNA methylation stabilizes gene silencing in embryonic tissues but is less important in extra-embryonic tissues, where PRC-mediated silencing might predominate. This mechanism of postzygotic gene silencing occurs in X-chromosome inactivation and in some forms of autosomal gene imprinting. H3K9 methylation is shown in green, H3K27 methylation in yellow, histone arginine methylation in pink and DNA methylation in red.

non-coding RNAs (*Kcnq1ot1* and *Xist*) begin to be expressed from the paternal allele in the two-cell embryo, and gene silencing *in cis* and the acquisition of histone modifications follow during the next few cleavage divisions and are largely complete by the blastocyst stage[34] (Fig. 3b).

The second model of how imprinted genes are silenced involves an epigenetically regulated chromatin insulator. In this model, tissue-specific enhancers are located on one side of the DMR overlapping with the insulator, whereas the silenced genes are on the other side[36]. Silencing occurs when the DMR is unmethylated and binds chromatin-organizing proteins such as CTCF (CCCTC-binding factor), resulting in a higher-order chromatin structure that prevents interactions between remote enhancers and promoters[37].

X-chromosome inactivation is another example of a relatively stable epigenetic silencing event; in this case, large regions of a whole chromosome are involved. In mice, imprinted X-chromosome inactivation is probably largely initiated by expression of *Xist* from the paternal chromosome at the two-cell stage[38]. (The nature of the imprinting leading to paternal expression is still unknown, but it is unlikely to be DNA methylation.) Imprinted X-chromosome inactivation is then stable (even in the absence of DNA methylation[39]) in the extra-embryonic tissues. Although the PcG protein system (which confers H3K27 methylation marks) has some influence on gene silencing, these modifications do not seem to confer heritable silencing[40]. Random X-chromosome inactivation is initiated in the epiblast after reprogramming of imprinted inactivation[41,42]. This reprogramming might be initiated by the silencing of *Xist* expression, and if this is the case, it is possible that the mitotic 'memory' for inactivation simply resides in the expression of *Xist*. The subsequent upregulation of *Xist* expression during the differentiation of epiblast cells is again followed by coating, gene silencing and acquisition of histone marks[43]. However, in contrast to imprinted X-chromosome inactivation, CpG islands in inactivated genes on the X chromosome become methylated and, although it has not been tested genetically, this might constitute long-term memory for inactivation during embryonic and adult life[43] (Fig. 3b). It is important to note that this methylation of CpG islands seems to be a dead end in that it does not need to be reprogrammed during the normal life cycle. (In the germ line, the inactivated X chromosome does not become methylated.)

Breaking stability by epigenetic reprogramming

DNA-methylation patterns that have been acquired during development are stable in somatic cells and during adult life. DNA-methylation patterns are somatically heritable essentially through the action of DNMT1, the maintenance methyltransferase[44]. At most CpG sites, the error rate of maintaining methylation (~1% per division) is low in relation to the number of cell divisions that are needed to produce a mammalian organism (44 for humans). Indeed, methylation of CpG islands is never erased during normal development. By contrast, methylation of CpG islands in imprinted-gene DMRs needs to be erased in the germ line so that gender-specific methylation can be imposed subsequently, during germ-cell development. This erasure takes place in a defined period — from E10.5 to E12.5 in PGCs — in all imprinted genes that have been tested[45,46], and it could occur by active demethylation of DNA by an unknown mechanism, possibly involving DNA repair (discussed later). This mechanism for erasure might also underlie the demethylation and activation of non-imprinted genes such as *Mvh*, *Dazl* and *Sycp3,* which takes place at about the same stage[20] (Fig. 4a).

Epigenetic reprogramming in PGCs entails widespread loss of DNA methylation, as well as H3K9 methylation[47]. In addition to the erasure of genomic imprints, this epigenetic reprogramming might also help to return PGCs to a pluripotent state (because at these stages of PGC development, pluripotent embryonic germ cells can be established in culture), through the reactivation of genes such as *Nanog*. Not all genomic methylation is lost, however, at these stages; some transposons such as IAPs remain fairly highly methylated[48]. Later in oogenesis and spermatogenesis, *de novo* methylation occurs not only sex-specifically in imprinted genes but also in transposons and in single-copy gene sequences. For example, the *Nanog* promoter becomes highly methylated in mature sperm[49].

Distinct genome-wide reprogramming events also occur immediately after fertilization and during early pre-implantation development (Fig. 4b). Many sequences in the paternal genome become suddenly demethylated shortly after fertilization[50–53]. This demethylation occurs after the removal of protamines (basic proteins that are associated with DNA in sperm) and the acquisition of histones by the paternal genome during the long G1 phase, before DNA replication. Methylation can be observed by staining cells with an immunofluorescently labelled antibody specific for 5-methylcytosine. Judged by the substantial loss of immunofluorescence signal, together with the considerable loss of methylation of *Line1* elements as determined by bisulphite sequencing[48], the paternal genome loses a significant amount of methylation, although more precise measurements and more information about which sequences are affected and unaffected would be valuable. Sequences that are known not to be affected include IAPs and paternally methylated DMRs in imprinted genes (Fig. 4c). A recent study provides intriguing insight into a protein that might protect the genome from demethylation. The protein stella (also known as DPPA3) binds to DNA and was originally identified because expression of the encoding gene is upregulated during early PGC development. Stella is present in large amounts in oocytes and, after fertilization, translocates to both pronuclei. Deletion of the gene from the oocyte (and therefore removal of the protein from the zygote) results in early pre-implantation lethality of embryos, as well as loss of methylation of the following sequences: the maternally methylated genes *Peg1* (also known as *Mest*), *Peg5* (also known as *Nnat*) and *Peg10*; the paternally methylated genes *H19* and *Rasgrf1* (Ras protein-specific guanine-nucleotide-releasing factor 1); and IAPs[54]. So stella might, either directly or indirectly, protect specific sequences from demethylation in the zygote, but it is unknown how other sequences are protected (Fig. 4c).

The mechanism of active demethylation in the zygote is still unknown. However, the DNA deaminases AID and APOBEC1 have been shown *in vitro* to deaminate 5-methylcytosine in DNA to thymine[55]; this results in T•G mismatches, which can be repaired by the base-excision repair pathway. Interestingly, *Aid* and *Apobec1* are located in a cluster of genes with *Stella*, growth differentiation factor 3 (*Gdf3*) and *Nanog*. *Stella*, *Gdf3* and *Nanog* are all expressed in pluripotent tissues, and *Gdf3* and *Nanog* have important roles in conferring stem-cell identity on ES cells. Indeed, *Aid* and *Apobec1* are also expressed by oocytes, stem cells and germ cells[55], and recent work shows that *in vivo* targeting of AID to the methylated *H19* DMR in the zygote results in efficient and substantial demethylation of this region (C. F. Chan, H. Morgan, F. Santos, D. Lucifero, S. Petersen-Mahrt, W. Dean and W.R., unpublished observations). Although it is unclear whether AID and/or APOBEC1 are responsible for the demethylation of the paternal genome in the zygote, the evidence suggests that base-excision or mismatch repair might have a role in this process. I think that this suggestion is supported by the recent identification of a DNA glycosylase–lyase — DEMETER — that preferentially excises 5-methylcytosine from DNA in *Arabidopsis thaliana*[56,57]. DEMETER is required for the demethylation and activation of the imprinted gene *MEDEA* (see page 418). Another DNA-damage-responsive gene, the mouse gene *Gadd45* (growth arrest and DNA-damage-inducible 45), might also have a role in demethylation[58].

Although there have been suggestions that the methyl group could be directly removed from DNA by hydrolytic attack or by oxidation, these mechanisms have not been substantiated[2]. The relative flexibility of histone methylation might be brought about by the attachment of the methyl group through a carbon–nitrogen bond, together with the existence of enzymes that can directly remove the methyl group, leaving the rest of the histone molecule intact[12]. By contrast, the current evidence suggests that methyl groups attach through a carbon–carbon bond to the cytosine base and therefore might not be able to be directly removed, so demethylation inevitably has to proceed by pathways that involve base-excision or mismatch repair[55–57].

This active demethylation of the paternal genome is followed by passive demethylation of both maternal and paternal genomes, presumably brought about by exclusion of DNMT1o (the main form of DNMT1 present in the oocyte) from the nuclei of pre-implantation embryos[27]. Although DNMT1s can maintain the methylation of imprinted-gene DMRs during this period, total genome methylation decreases, reaching an overall low at the blastocyst stage. The purpose of active and passive demethylation during early embryogenesis is unknown. Demethylation of the paternal genome has been proposed to account for the paucity of paternal imprints[59] or to be a consequence of DNA-repair processes that are potentially involved in the protamine-to-histone transition[53]. General demethylation during this period could also have a role in returning the gametic genomes to pluripotency. For example, early expression of genes such as *Oct4* or *Nanog* is required for the establishment and maintenance of the inner-cell-mass lineage in the blastocyst[60]. Because the *Nanog* and *Oct4* promoters are methylated in sperm, and because methylation of these promoters is repressive, they need to be demethylated for proper expression to occur (Fig. 4b).

Epigenetic spillover across generations

Many of the epigenetic marks that are inherited and acquired by germ cells are therefore erased in PGCs and in early embryos, making way for new generations to develop and grow into adults purely on the basis of their genetic make-up. However, it also seems that epigenetic information can spill over to the next generation. The ability of somatic cells in the offspring to inherit the methylation of imprinted genes from parental germ cells is a mechanistic example of this (Fig. 4c). Another important example of spillover is inheritance of the epigenetic states conferred on some genes by adjacent insertion of IAPs. This can alter the expression of the endogenous genes; however, more importantly, the epigenetic state of the IAP (that is, methylated or unmethylated) regulates the expression of the nearby gene[61]. Because IAPs seem generally resistant to reprogramming during PGC and pre-implantation development, the state of expression of the genes that are regulated by IAP insertion can be inherited across several generations. It is interesting to note that there is an example of epigenetic inheritance being maternally transmitted but not paternally transmitted (the agouti viable yellow epiallele in mice), and the methylation of the IAP in the sperm is, unusually, erased in the zygote in this case[62]. So epigenetic inheritance is 'broken' by erasure of methylation of the paternal genome after fertilization.

There are other possible spillovers across generations. In *Caenorhabditis elegans*, the X chromosomes are epigenetically marked (by histone modifications) during gametogenesis, and some of these marks are maintained for several cell divisions in the new embryo (for an unknown reason)[63]. In mammalian embryos, some of the histone modifications acquired during the silencing of X-linked genes in spermatogenesis might be carried over into the zygote, leading to early silencing of some genes on the paternal X chromosome without the action of *Xist*[64].

One other area that is unique to mammalian biology deserves consideration with regard to epigenetic spillovers from the previous generation. At present, we have no understanding of how molecular decisions are taken to set up the first two cell lineages in the embryo: the trophectoderm and the inner cell mass[65]. However, a recent study suggests that differential histone arginine methylation of individual blastomeres, as early as the four-cell stage, could be one of the earliest marks for this lineage commitment[66]. There is much work to be done in this area, but it is an exciting possibility that the spillover of epigenetic marks from the gametes of parents might be responsible for setting up some of the earliest developmental decisions in the newly developing embryo.

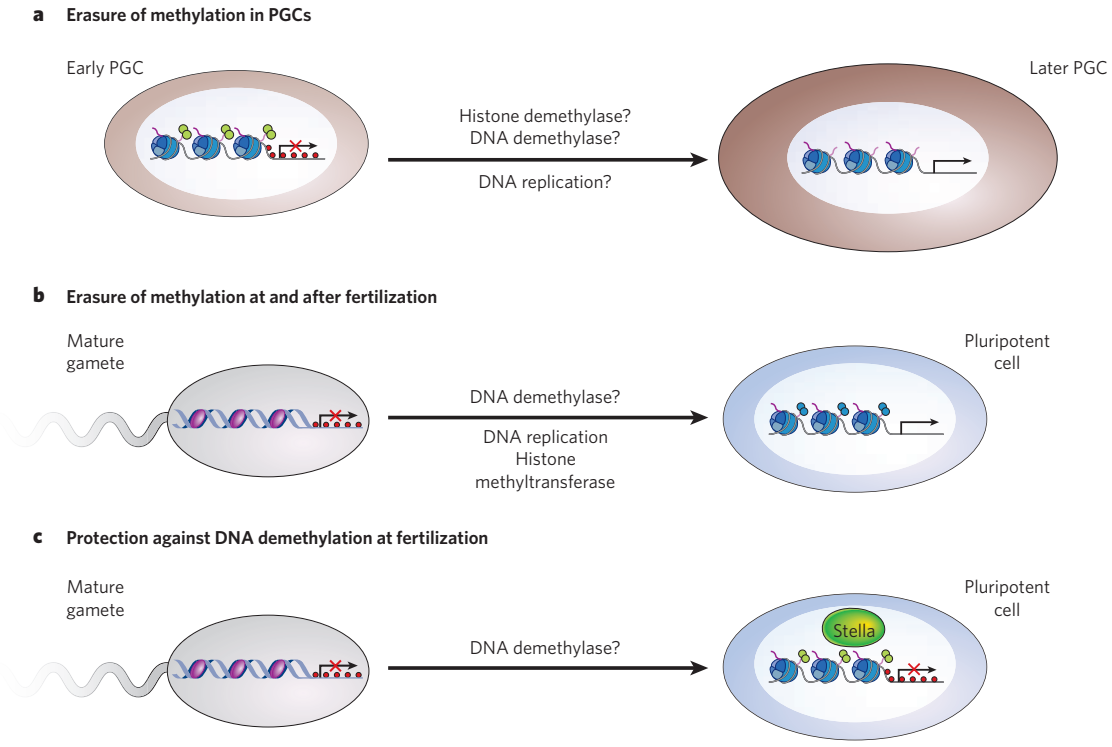


Figure 4 | Reprogramming of epigenetic marks in the germ line and the early embryo. a, During the development of PGCs, methylation of CpG islands in imprinted genes and other genes can be erased. This is a rapid process and might involve a demethylase or might occur by DNA replication without methylation being maintained. **b**, Many gene sequences that are methylated in mature gametes become demethylated in the early embryo. Some of this demethylation occurs in the absence of DNA replication and is therefore likely to be mediated by a demethylase. Demethylation might be important for the expression of pluripotency-associated genes. Active histone marks are also likely to be important for the expression of pluripotency-associated genes. **c**, In mature gametes, some DNA sequences that are methylated are protected from demethylation at or after fertilization. These sequences include imprinted genes and some transposons. The protein stella has recently been implicated in protection against demethylation at fertilization. H3K9 methylation is shown in green, H3K4 methylation in blue and DNA methylation in red.

Conclusions and outlook

Development might be a one-way street because of the somatic inheritance of epigenetic marks. Whether there is a linear relationship between acquisition of epigenetic marks and developmental progression is doubtful; some key restrictions in developmental potential that are brought about by epigenetic regulation might occur very early in development. Judging from somatic-cell nuclear-transfer experiments, it is far from clear whether more-differentiated cells have more epigenetic marks or have marks that are more difficult for the oocyte to reprogramme[67].

Natural epigenetic reprogramming might be needed to ensure that development can start afresh in every new generation. Although various mechanisms for the rapid erasure of histone modifications have recently been identified, the mechanism of DNA demethylation still needs to be determined. Recent work on the erasure of DNA methylation from imprinted plant genes shows that base-excision repair has an important role, and it is possible that this is also the case in mammals. Because of the generally accurate heritability of DNA methylation and because of its chemical stability, erasure of DNA methylation might only be possible either by replicating DNA in the absence of DNMT1 or by breaking DNA.

It is fascinating to see that both transcription-factor interactions and epigenetic programming and reprogramming seem to be needed to maintain pluripotency in early embryos and ES cells. Indeed, experimental reprogramming of differentiated nuclei without using somatic-cell nuclear transfer or cell fusion has been achieved recently, using a mix of pluripotency factors[1]. It could be expected that forcing the expression of pluripotency transcription-factor networks would also activate epigenetic reprogramming factors, but whether this occurs is unclear. Perhaps combinations of transcription factors and epigenetic reprogramming factors are needed for more complete reprogramming of somatic cells to a pluripotent state, and this would be of great fundamental scientific and medical interest.

In the animal kingdom, some epigenetic systems, such as imprinting, have evolved only in mammals. Many of the basic molecular building blocks for epigenetics, such as the enzymes for DNA methylation and histone modifications, are highly conserved in vertebrates, but the regulation of epigenetic modifiers might evolve more rapidly together with specific developmental strategies. Therefore, evolutionary epigenetics and epigenomics will have an important role in discovering links between developmental adaptations and epigenetic regulators.

There is probably a conflict between the requirement for erasing epigenetic marks between generations and the requirement for maintaining others, such as those in imprinted genes and in some transposons. This conflict most probably underlies the observation that some epigenetic marks are not erased between generations, thereby leading to multigenerational influences on inheritance and phenotype (see page 396). Epigenetic inheritance across generations is relatively common in plants, but it is still unclear how widespread this phenomenon is in mammals or whether it has any role in shaping evolution[61].

An exciting question for future work is whether segregation of epigenetic marks in early development has any primary role in determining cell and lineage commitment. For example, the mechanism by which the first two cell lineages are allocated in mammalian pre-implantation embryos, although a matter of hot debate, is not really understood. An epigenetic hypothesis might allow us to take a fresh look at a long-standing fundamental problem in developmental biology. ■

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14**, R47–R58 (2005).
3. Allis, C. D., Jenuwein, T. & Reinberg, D. (eds) *Epigenetics* (Cold Spring Harbor Laboratory Press, Woodbury, 2007).
4. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
5. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Rev. Genet.* **3**, 662–673 (2002).
6. Turner, B. M. Defining an epigenetic code. *Nature Cell Biol.* **9**, 2–6 (2007).
7. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).
8. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
9. Szutorisz, H. *et al.* Formation of an active tissue-specific chromatin domain initiated by epigenetic marking at the embryonic stem cell stage. *Mol. Cell. Biol.* **25**, 1804–1820 (2005).
10. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature Cell Biol.* **8**, 532–538 (2006).
11. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
12. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nature Rev. Genet.* **7**, 715–727 (2006).
13. Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genet.* **39**, 237–242 (2007).
14. Feldman, N. Y. *et al.* G9a-mediated irreversible epigenetic inactivation of *Oct-3/4* during early embryogenesis. *Nature Cell Biol.* **8**, 188–194 (2006).
15. Simpson, A. J., Caballero, O. L., Jungbluth, A., Chen, Y. T. & Old, L. J. Cancer/testis antigens, gametogenesis and cancer. *Nature Rev. Cancer* **5**, 615–625 (2005).
16. Hochedlinger, K., Yamada, Y., Beard, C. & Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121,** 465–477 (2005).
17. Boiani, M., Eckardt, S., Scholer, H. R. & McLaughlin, K. J. Oct4 distribution and level in mouse clones: consequences for pluripotency. *Genes Dev.* **16**, 1209–1219 (2002).
18. Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
19. Ancelin, K. *et al.* Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nature Cell Biol.* **8**, 623–630 (2006).
20. Maatouk, D. M. *et al.* DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**, 3411–3418 (2006).
21. Surani, A. & Reik, W. in *Epigenetics* (eds Allis, C. D., Jenuwein, T. & Reinberg, D.) 315–327 (Cold Spring Harbor Laboratory Press, Woodbury, 2007).
22. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96–99 (2004).
23. Barlow, D. P. Methylation and imprinting: from host defense to gene regulation? *Science* **260**, 309–310 (1993).
24. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539 (2001).
25. Kaneda, M. *et al.* Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900–903 (2004).
26. Jelinic, P., Stehle, J. C. & Shaw, P. The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in *H19* imprinting control region methylation. *PLoS Biol.* [online] **4**, e355 (2006) (doi:10.1371/journal.pbio.0040355).
27. Howell, C. Y. *et al.* Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene. *Cell* **104**, 829–838 (2001).
28. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
29. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002)
30. Mancini-Dinardo, D., Steele, S. J., Levorse, J. M., Ingram, R. S. & Tilghman, S. M. Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* **20**, 1268–1282 (2006).
31. Lewis, A. *et al.* Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nature Genet.* **36**, 1291–1295 (2004).
32. Umlauf, D. *et al.* Imprinting along the *Kcnq1* domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nature Genet.* **36**, 1296–1300 (2004).
33. Kanduri, C., Thakur, N. & Pandey, R. R. The length of the transcript encoded from the *Kcnq1ot1* antisense promoter determines the degree of silencing. *EMBO J.* **25**, 2096–2106 (2006).
34. Lewis, A. *et al.* Epigenetic dynamics of the *Kcnq1* imprinted domain in the early embryo. *Development* **133**, 4203–4210 (2006).
35. Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for *Xist* RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **20**, 2223–2227 (2006).
36. Verona, R. I., Mann, M. R. & Bartolomei, M. S. Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annu. Rev. Cell Dev. Biol.***19**, 237–259 (2003).
37. Kurukuti, S. *et al.* CTCF binding at the *H19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proc. Natl Acad. Sci. USA* **103**, 10684–10689 (2006).
38. Okamoto, I. *et al.* Evidence for *de novo* imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature* **438**, 369–373 (2005).
39. Sado, T. *et al.* X inactivation in the mouse embryo deficient for *Dnmt1*: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev. Biol.* **225**, 294–303 (2000).
40. Kohlmaier, A. *et al.* A chromosomal memory triggered by *Xist* regulates histone methylation in X inactivation. *PLoS Biol.* [online] **2,** e171 (2004) (doi:10.1371/journal.pbio.0020171).
41. Mak, W. *et al.* Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303**, 666–669 (2004).
42. Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. & Heard, E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644–649 (2004).
43. Heard, E. & Disteche, C. M. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* **20**, 1848–1867 (2006).
44. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
45. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**, 15–23 (2002).
46. Lee, J. *et al.* Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* **129**, 1807–1817 (2002).
47. Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* **278**, 440–458 (2005).

48. Lane, N. *et al*. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35**, 88–93 (2003).
49. Imamura, M. *et al*. Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* [online] **6**, 34 (2006) (doi:10.1186/1471-213X-6-34).
50. Oswald, J. *et al*. Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478 (2000).
51. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
52. Dean, W. *et al*. Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc. Natl Acad. Sci. USA* **98**, 13734–13738 (2001).
53. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* **241**, 172–182 (2002).
54. Nakamura, T. *et al*. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biol.* **9**, 64–71 (2006).
55. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279**, 52353–52360 (2004).
56. Gehring, M. *et al*. DEMETER DNA glycosylase establishes *MEDEA* Polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**, 495–506 (2006).
57. Morales-Ruiz, T. *et al. DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. *Proc. Natl Acad. Sci. USA* **103**, 6853–6858 (2006).
58. Barreto, G. *et al.* Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**, 671–675 (2007).
59. Reik, W. & Walter, J. Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. *Nature Genet.* **27**, 255–256 (2001).
60. Smith, A. G. Embryo-derived stem cells: of mice and men. *Annu. Rev. Cell Dev. Biol.* **17**, 435–462 (2002).
61. Whitelaw, N. C. & Whitelaw, E. How lifetimes shape epigenotype within and across generations. *Hum. Mol. Genet.* **15**, R131–R137 (2006).
62. Blewitt, M. E., Vickaryous, N. K., Paldi, A., Koseki, H. & Whitelaw, E. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet.* [online] **2**, e49 (2006) (doi:10.1371/journal.pgen.0020049).
63. Bean, C. J., Schaner, C. E. & Kelly, W. G. Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. *Nature Genet.* **36**, 100–105 (2004).
64. Namekawa, S. H. *et al*. Postmeiotic sex chromatin in the male germline of mice. *Curr. Biol.* **16**, 660–667 (2006).
65. Rossant, J. Lineage development and polar asymmetries in the peri-implantation mouse blastocyst. *Semin. Cell Dev. Biol.* **15**, 573–581 (2004).
66. Torres-Padilla, M.E., Parfitt, D.E., Kouzarides, T. & Zernicka-Goetz, M. Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* **445**, 214–218 (2007).
67. Yang, X. *et al.* Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. *Nature Genet.* **39**, 295–302 (2007).

Acknowledgements I thank all my colleagues, past and present, for their contributions to the work and ideas described in this paper, especially W. Dean, F. Santos, A. Lewis, and G. Smits. Funding from the Biotechnology and Biological Sciences Research Council, the Medical Research Council, the European Union Epigenome Network of Excellence, CellCentric and the Department of Trade & Industry is gratefully acknowledged.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to the author (wolf.reik@bbsrc.ac.uk).

Phenotypic plasticity and the epigenetics of human disease

Andrew P. Feinberg[1]

It is becoming clear that epigenetic changes are involved in human disease as well as during normal development. A unifying theme of disease epigenetics is defects in phenotypic plasticity — cells' ability to change their behaviour in response to internal or external environmental cues. This model proposes that hereditary disorders of the epigenetic apparatus lead to developmental defects, that cancer epigenetics involves disruption of the stem-cell programme, and that common diseases with late-onset phenotypes involve interactions between the epigenome, the genome and the environment. Increased understanding of epigenetic-disease mechanisms could lead to disease-risk stratification for targeted intervention and to targeted therapies.

The original definition of epigenetics by Waddington in 1942 (ref. 1) — the idea that that phenotype arises from genotype through programmed change — is now what is considered to be developmental biology. The modern definition of epigenetics is information heritable during cell division other than the DNA sequence itself. It is becoming increasingly clear that there is great overlap between these two definitions — developmental processes are regulated largely by epigenetics, because different cell types maintain their fate during cell division even though their DNA sequences are essentially the same.

What is epigenetic disease? Genetic lesions — sequence changes, breakages and deletions — can be easily visualized, but what about epigenetic lesions? Several defects in the epigenome are known to lead to disease (Fig. 1), including changes in the localized or global density of DNA methylation, and incorrect histone modification. Other defects that might cause disease involve altered distribution or function of chromatin-modifying proteins that, in turn, leads to aberrant gene expression. Another intriguing possibility is the disruption of higher-order loop structure in disease (Fig. 1).

Studies of monogenic disorders involving imprinted genes or the epigenetic machinery have revealed a great deal about the nature of the *cis*-acting regulatory marks and *trans*-acting factors that modulate the epigenome. Two decades ago, cancer epigenetics was viewed with some scepticism, but it is now widely accepted. However, important questions about the mechanism, timing and consequences of epigenetic disruption remain. Whether other common diseases have an epigenetic basis is still open to speculation, but if they do, this holds great promise for medicine.

Here I review the epigenetics of single-gene disorders, cancer and common complex diseases. I suggest that a common theme to disease epigenetics is the disruption of phenotypic plasticity — the ability of cells to change their behaviour in response to internal or external environmental cues — an idea that resonates with Waddington's original definition of epigenetics. I also discuss therapeutic implications of disease epigenetics.

Imprinted gene disorders

There are two classes of monogenic epigenetic disease: those involving genes that are regulated epigenetically, such as imprinted genes, and those that affect the epigenome as a whole, such as modifiers of DNA methylation.

Epigenetic disease genes

An example of the first class of monogenic epigenetic disease is Beckwith–Wiedemann syndrome, which is characterized by prenatal overgrowth, a midline abdominal wall and other malformations, and cancer. Studies of patients with Beckwith–Wiedemann syndrome have taught us a great deal about the mechanisms of normal imprinting. Patients with Beckwith–Wiedemann syndrome show disrupted imprinting of either or both of two neighbouring imprinted subdomains on 11p15, revealing clustering of imprinted genes (Fig. 2). The first is the *H19*/*IGF2* (imprinted, maternally expressed, untranslated mRNA/insulin-like growth factor 2) imprinted subdomain, which is regulated by a differentially methylated region (DMR) that is methylated on the paternal but not the maternal allele. The second subdomain includes $p57^{KIP2}$ (a cyclin-dependent kinase inhibitor), *TSSC3* (a pleckstrin homology-like domain), *SLC22A1* (an organic cation transporter), K_vLQT1 (a voltage-gated potassium channel) and *LIT1* (*KCNQ1* overlapping transcript 1), with the subdomain regulated by a second DMR, upstream of *LIT1*, that is normally methylated on the maternal but not the paternal allele (see ref. 2 for a review). In patients with Beckwith–Wiedemann syndrome, microdeletions within each imprinted subdomain have confirmed the regulatory role of these sequences, because individuals with these deletions show loss of normal imprinted gene regulation[3,4] (Fig. 2). Some patients with Beckwith–Wiedemann syndrome show loss of imprinting of *IGF2*, which leads to a double dose of this autocrine factor, resulting in tissue overgrowth and increased cancer risk[5]. The mechanism involves aberrant methylation of the maternal *H19* DMR (Fig. 2). Other patients with this syndrome show localized abnormalities of allele-specific chromatin modification[6] affecting $p57^{KIP2}$, a cyclin-dependent kinase inhibitor (Fig. 2). Thus, Beckwith–Wiedemann syndrome illustrates hierarchical organization of epigenetic regulation in progressively larger domains.

A pair of imprinted-gene disorders that are associated with mental retardation — Prader–Willi syndrome and Angelman syndrome — involve adjacent reciprocally imprinted genes, *SNRPN* (small nuclear ribonucleoprotein polypeptide N) and *UBE3A* (ubiquitin–protein ligase E3A), on chromosome 15. Microdeletions in patients with both Prader–Willi syndrome and Angelman syndrome reveal the location of the domain that regulates imprinting of both genes[7].

Another pair of human disorders caused by different alterations at the same locus are Albright hereditary osteodystrophy and

[1]Department of Medicine and Center for Epigenetics, Institute for Basic Biomedical Sciences, Johns Hopkins University School of Medicine, 720 Rutland Avenue, Baltimore, Maryland 21205, USA.

pseudohypoparathyroidism type IA (PHPIA). Albright hereditary osteodystrophy is characterized by short stature and ectopic calcifications, and caused by mutational inactivation of guanine nucleotide regulatory protein (encoded by *GNAS1*). PHPIA is a more severe phenotype of multiple hormone resistance caused by tissue-specific differential imprinting of splice variants of the same gene[8]. It is unlikely that this complex pattern of imprinting would have been understood without concomitant clinical studies.

Single-gene disorders of the epigenetic machinery

The other class of monogenic epigenetic disease involves genes that encode components of the machinery that regulates the epigenome. Mutation of these genes causes developmental disorders. For example, Rett syndrome involves mutations of the methyl CpG-binding protein 2 (*MeCP2*) gene, which encodes a protein that binds to methylated DNA sequences[9]. In Rett syndrome, DNA methylation proceeds normally but epigenetic silencing is impaired because of a failure to properly recognize this mark[10] (Fig. 3). What is striking about the phenotype of this disorder is that prenatal and early infant development is normal, and erosion of neurodevelopmental milestones is not seen until later childhood.

Epigenetically disrupted development can occur in various biological pathways or systems. Immunodeficiency/centromeric instability/facial anomalies (ICF) syndrome, for example, affects the immune system and involves mutations of the DNA methyltransferase gene *DNMT3B*, which is responsible for *de novo* DNA methylation during development[11]. Patients wih ICF syndrome show failure of normal immune system development as well as developmental dysmorphology, which could involve failure of heterochromatin formation and might result from *DNMT3B* having a role in immunoglobulin gene silencing and reactivation[12].

A striking example of developmental disruption caused by mutations in a chromatin factor gene is alpha-thalassaemia/mental retardation, X-linked (ATRX) syndrome, the gene for which is a helicase involved in chromatin remodelling. Mutations lead to defects in psychomotor, urogenital and haematopoietic development, with maturational defects in erythroid precursors resembling those of alpha-thalassaemia[13]. Rubinstein–Taybi syndrome involves the CREB (cyclic-AMP responsive-element-binding protein)-binding protein CBP, which has histone acetyltransferase activity, and mutations in *CBP* lead to skeletal and cardiac malformations, as well as neurodevelopmental malformations and loss of neural plasticity[14]. A common theme of these disorders is that mutations in epigenome regulators cause developmental disruption and often cause phenotypic changes in multiple organ systems.

DNA methylation in cancer

Cancer is commonly characterized as showing global hypomethylation and site-specific gene hypermethylation, but a more accurate description is that cancer involves both global and gene-specific hypomethylation and hypermethylation, as well as widespread chromatin modifications (Fig. 3). The first epigenetic change described in tumours was gene hypomethylation[15], and we now know that many growth-promoting genes are activated through hypomethylation in tumours, including *HRAS*, cyclin D2 and maspin in gastric cancer, carbonic anhydrase IX in

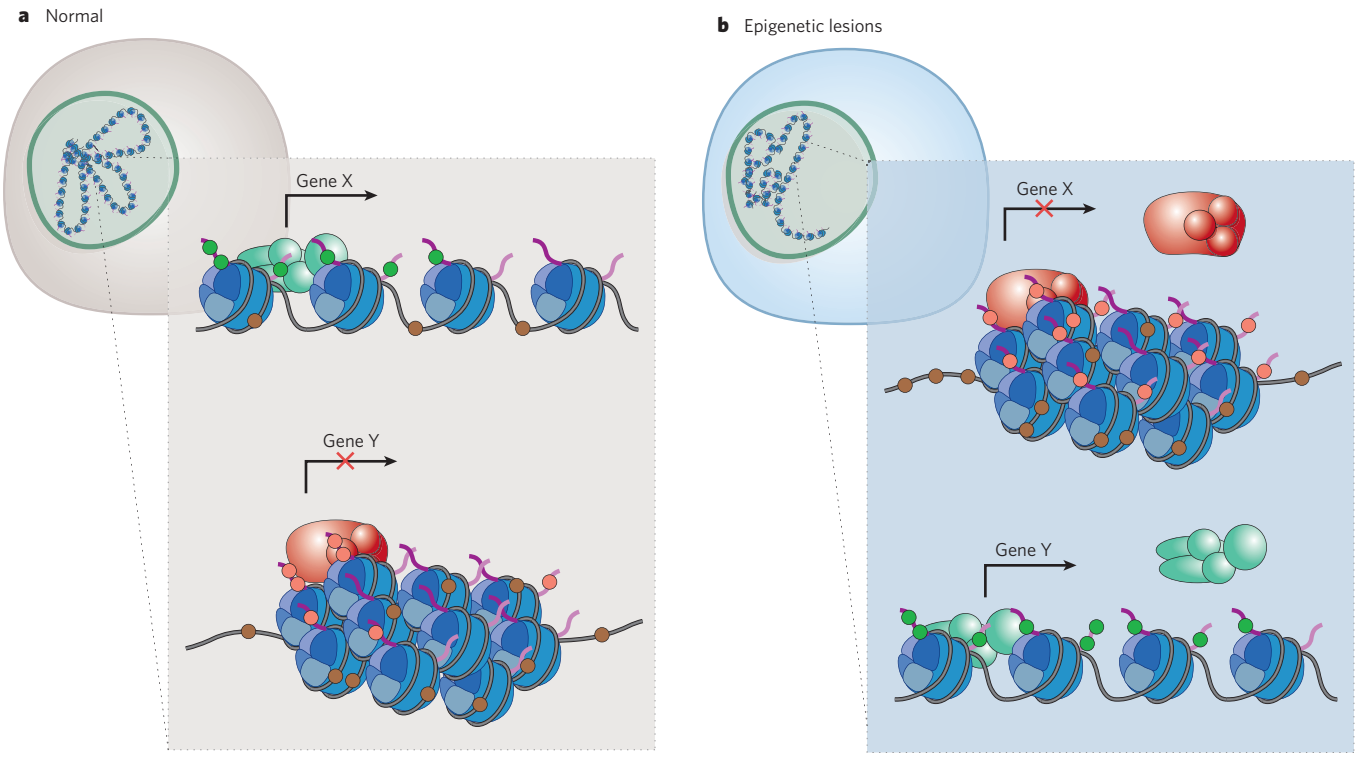


Figure 1 | The nature of epigenetic lesions. Although the nature of genetic lesions is well understood, epigenetic lesions have been more difficult to define. Here we depict known and possible defects in the epigenome that could lead to disease. **a**, X is a transcriptionally active gene with sparse DNA methylation (brown circles), an open chromatin structure, interaction with euchromatin proteins (green protein complex) and histone modifications such as H3K9 acetylation and H3K4 methylation (green circles). Y is a transcriptionally silent gene with dense DNA methylation, a closed chromatin structure, interaction with heterochromatin proteins (red protein complex) and histone modifications such as H3K27 methylation (pink circles). **b**, The abnormal cell could switch its epigenotype through the silencing of normally active genes or activation of normally silent genes, with the attendant changes in DNA methylation, histone modification and chromatin proteins. In addition, the epigenetic lesion could include a change in the number or density of heterochromatin proteins in gene X (such as EZH2 in cancer) or euchromatic proteins in gene Y (such as trithorax in leukaemia). There may also be an abnormally dense pattern of methylation in gene promoters (shown in gene X), and an overall reduction in DNA methylation (shown in gene Y) in cancer. The insets show that the higher-order loop configuration may be altered, although such structures are currently only beginning to be understood.

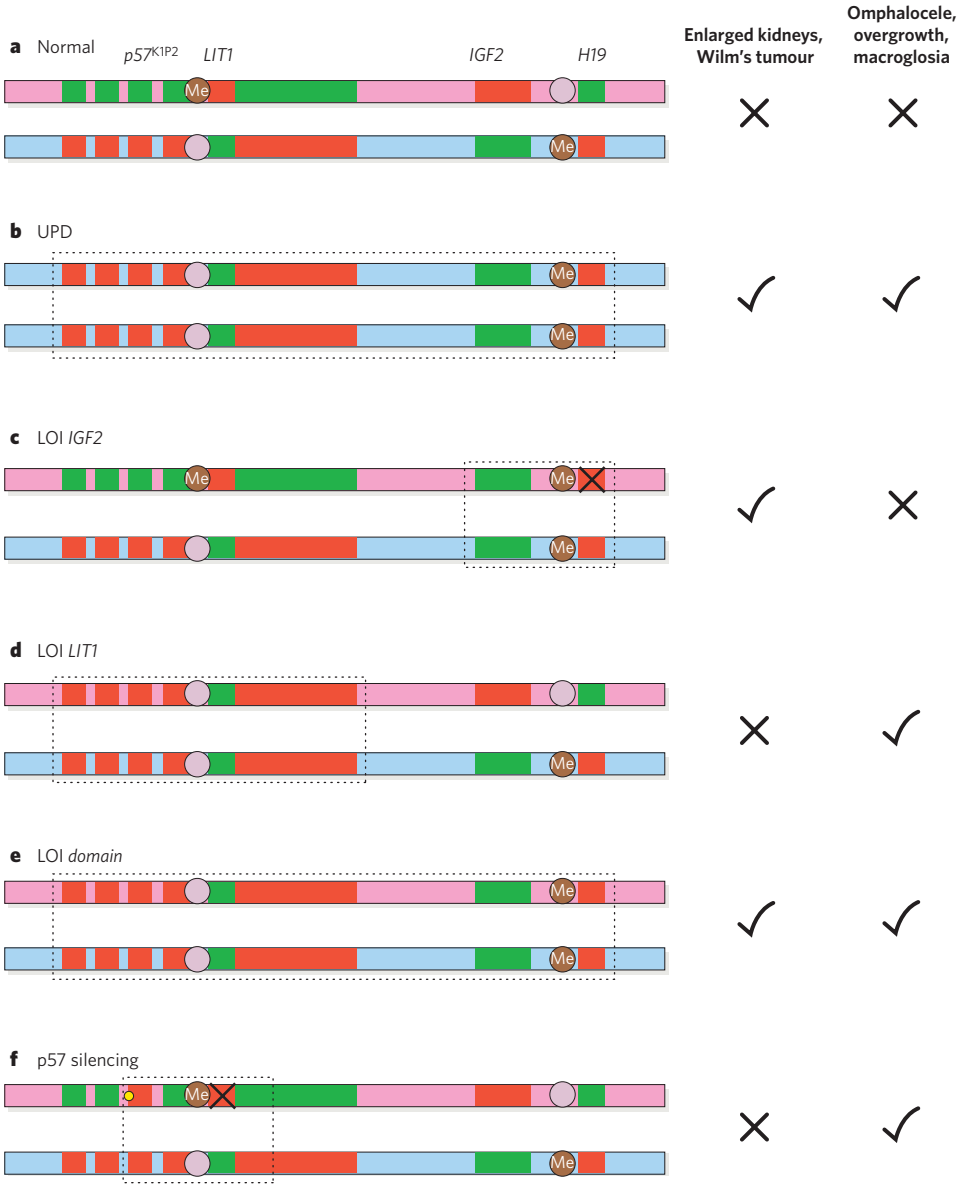


Figure 2 | Beckwith–Wiedemann syndrome as an example of a monogenic disease that reveals mechanisms of normal epigenetic regulation. Depicted are a pair of normal chromosomes (**a**) and several illustrative lesions (**b**–**f**); maternal chromosomes are pink and paternal blue, and DMRs are indicated (Me). Imprinted genes are depicted, not to scale or in their entirety, with green representing active and red representing silent alleles, respectively. Patients with uniparental disomy (UPD, **b**) have complete genetic replacement of the maternal allele region with a second paternal copy (dashed enclosure). Loss of imprinting (LOI) of *IGF2* (**c**) causes a switch in epigenotype of the *IGF2/H19* subdomain (dashed enclosure). LOI of *LIT1* (**d**) causes a switch in epigenotype of the *p57/K$_V$LQT1/LIT1* subdomain. Some patients show LOI of the entire imprinted gene domain in the absence of UPD (**e**). Other patients show localized chromatin disruption (small yellow circle, **f**) silencing *p57*^{KIP2}. Thus, imprinting is organized hierarchically into a large domain containing two smaller subdomains. In addition, some patients show microdeletions in either of the two domains (black crosses), revealing the location of imprinting control centres. The domain organization similarly reveals the contiguous gene syndrome nature of the disease. Patients with involvement (genetic or epigenetic) of the *IGF2/H19* domain have enlarged kidneys and Wilms' tumours. Patients with involvement of the *p57/K$_V$LQT1/LIT1* domain show somatic overgrowth, an enlarged tongue and omphalocele (in which abdominal organs protrude from the navel). And children with involvement of both domains show both phenotypes.

renal-cell cancer, and S100 calcium-binding protein A4 in colon cancer (see refs 16, 17 for reviews). In addition, many C/T (cancer/testis) genes that are expressed normally in the healthy testis are activated in other cells by hypomethylation in cancer, including the melanoma-associated antigen (*MAGE*) gene family, which has antigenic and immunotherapeutic value in melanoma and glioblastoma[16,17] and the oncogenic micro RNA let-7a-3 (ref. 18). Activation of the human papilloma virus HPV16 by hypomethylation is a major mechanism affecting tumour latency in cervical cancer[16,17]. Recently, oestrogen- and tamoxifen-induced activation of *PAX2* and endometrial proliferation, leading to cell proliferation, was found to be cancer-specific because of *PAX2* promoter hypomethylation in the tumours[19].

By contrast, tumour suppressor gene silencing has been linked to promoter hypermethylation, first described for *RB*, the gene associated with retinoblastoma[20], and many other tumour suppressor genes, including *p16*, *VHL* (von Hippel–Lindau), *MLH1*, *APC* (adenomatosis polyposis coli) and E-cadherin (see ref. 21 for a review). Recent high-throughput approaches have been used to identify other candidate genes[22,23]. An exciting demonstration of domain-wide silencing involves an entire chromosomal band[24], suggesting a disturbance of higher-order chromatin (Fig. 1). However, studies focused on loss of DNA methylation in cancer may have overlooked hypomethylation of tissue-specific methylation marks at CpG islands[25]. Indeed, whole-genome analysis suggests that CpG island hypermethylation may be less widespread than had been suspected[26]. It should also be noted that as many genes are silenced as are activated in tumours by both drug-induced hypomethylation and by knockdown of DNA methyltransferases[27], thus both hypomethylation and hypermethylation can lead to gene activation and gene silencing in cancer.

Loss of imprinting in cancer

The earliest clue that genomic imprinting might be involved in cancer came from two rare types of tumour: hydatidiform moles, which are malignant trophoblastic tumours caused by a pregnancy arising from two complete sets of the paternal genome, and ovarian teratomas, which are benign tumours with many tissue types that arise from two complete sets of the maternal genome. Molecular evidence for a role of genomic imprinting in cancer emerged from studies showing a universal loss of the maternal allele in Wilms' tumours and embryonal rhabdomyosarcoma, with loss of heterozygosity (LOH) of 11p15, implying that normally only the maternal allele of an as yet unidentified tumour suppressor gene might be expressed[28]. So it was surprising that the first molecular evidence for a role of genomic imprinting in cancer was loss of imprinting (LOI), causing abnormal activation of the normally silent copy of *IGF2*, an important autocrine growth factor, leading to pathological biallelic expression of *IGF2* in Wilms' tumours, the most common childhood solid tumour[29,30]. LOI refers either to aberrant activation of the

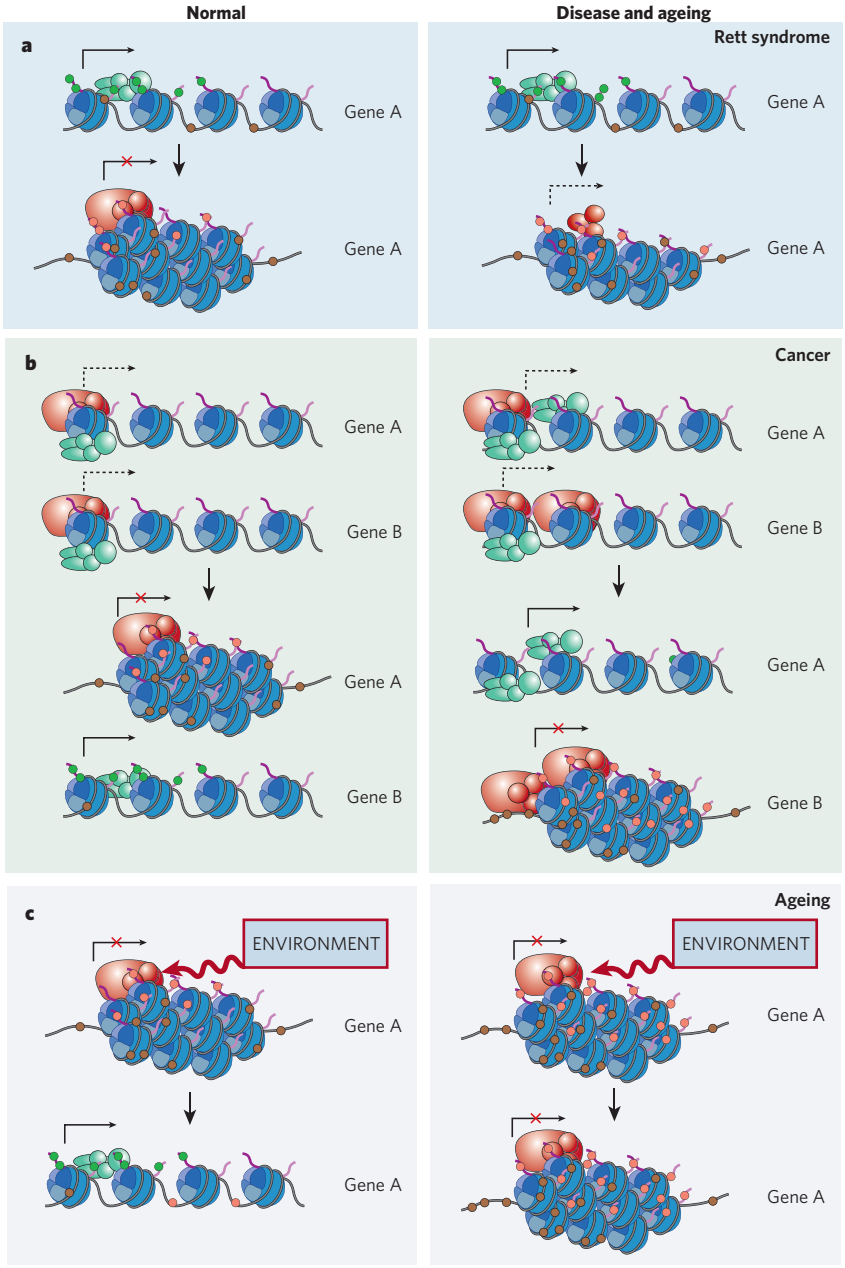


Figure 3 | Phenotypic plasticity and the epigenetics of human disease and ageing. A common feature of epigenetic lesions in human disease is that they affect a cell's ability to change its phenotype. **a**, In monogenic disorders such as Rett syndrome, a defect in the normal epigenetic apparatus itself impedes normal development. DNA methylation (brown circles on the DNA) proceeds normally but is not recognized owing to the absence of the MeCP2-methylation-interacting protein (large red oval). This leads to failure to completely silence genes appropriately during development (dashed arrow). **b**, Cancer involves many epigenetic lesions that could affect a pluripotent programme in tissue-specific stem cells, possibly leading to an incorrect distribution of differentiated cell lineages (indicated by the bivalent euchromatin and heterochromatin proteins shown in the upper left panel) and normal tissue-specific silencing of gene A and activation of gene B after differentiation (lower left panel). Examples of epigenetic lesions found in cancer include changes in chromatin proteins in stem cells caused by increased expression of *MLL1* in leukaemia (upper right panel, green complex above gene A representing *HOX* genes), leading to aberrant *HOX* expression in differentiated cancer lineages (lower right panel). Another epigenetic lesion found in cancer is increased expression of *EZH2* (upper right panel, red complex above gene B, representing diverse tumour suppressor genes), leading to aberrant silencing of these genes in differentiated cancer lineages (lower right panel). **c**, Ageing involves a loss of the normal plasticity of response to internal and external environmental signals. The epigenome could have an important role in ageing if the aged epigenome is less responsive to such signals. For example, a gene (at this point hypothetically) showing increased H3K9-methylation (upper right panel, red circles on nucleosomes) or DNA methylation (brown circles on DNA), might be relatively refractory to environmentally induced activation (lower right panel) than if the gene had not undergone age-dependent epigenetic change (left panels).

normally silent allele of an imprinted growth-promoting gene, or aberrant silencing of the normally expressed copy of an imprinted tumour suppressor gene, such as the still unidentified locus on 11p15 (ref. 31). Subsequently, LOI of *IGF2* has been found to be common in lung cancer[32], breast cancer[33], ovarian cancer[34] and glioma[35]. LOI of other genes include *ARHI* in breast cancer[36], *DLK1/GTL2* in pheochromocytoma, neuroblastoma and Wilms' tumour[37], and *PEG1* (also known as *MEST*) in breast cancer[38].

Chromatin and cancer

It has become increasingly clear that in cancer chromatin modifications are at least as widespread and important as alterations in DNA methylation. For example, overexpression of the polycomb group protein EZH2, a H3 lysine-27 (H3K27) histone methyltransferase, is found in metastatic prostate cancer and may lead to widespread transcriptional repression[39] (Fig. 3). Generalized loss of H4 acetylated Lys-16 (H4K16ac) and trimethylated Lys-20 (H4K20me3) is found in lymphoma and colorectal cancer, which could also lead to transcriptional silencing[40] (Fig. 3). It is not surprising that both DNA methylation and histone modification are altered in cancer, given their interdependence in normal development. For example, in the fungus *Neurospora* DNA methylation depends on H3K9 methylation[41], and in mice DNA methylation of homeobox (*Hox*) genes depends on a full length *Mll* (myeloid/lymphoid or mixed-lineage leukaemia) gene[42]. DNMT1 interacts with the H3K9 methyltransferases G9a and SUV39H1, which are needed for normal replication-dependent DNA methylation[43]. Some chromosomal rearrangements and, less commonly, mutations in cancer act by causing widespread chromatin disruption. *MLL1*, which is rearranged and activated in acute lymphoblastic leukaemia, methylates H3K4 to activate gene expression and interacts with integrase integrator 1 (*INI1*) in the SWI/SNF chromatin remodelling complex[44]. *INI1* is mutated in rhabdoid tumour, a deadly soft-tissue malignancy[45]. Sotos syndrome, which is characterized by tissue overgrowth, leukaemia and Wilms' tumour, is caused by mutations in *NSD1*, an H3K36/H4K20 methyltransferase[46]. Thus, a strong argument can be made for chromatin modifications driving epigenetic disruptions during cancer development.

The argument for causality

One problem with the idea that alterations in DNA methylation underlie cancer is that no mutations in either the methylation modification or

the recognition machinery have yet been identified in human cancer. Indeed, congenital disorders involving such modifications — for example, Rett syndrome (*MeCP2*) and ICF syndrome (*DNMT3B*) — do not carry an increased cancer risk, in contrast to the chromatin-modifying disruptions described above. In addition, epigenetic changes might be a consequence of altered gene expression rather than causal; it has been known since the 1980s that numerous genes are aberrantly expressed in tumours[47]. Furthermore, activation of tumour suppressor genes by 5-aza-2′-deoxycytidine or *DNMT1* knockout may not be stable, as has been shown for both *MLH1* (ref. 48) and *p16* (ref. 49), suggesting that the altered methylation might be a consequence rather than a cause of gene silencing.

So how can a convincing causal argument be made? Good evidence would be constitutional epigenetic alterations linked to cancer risk. The first such example was Beckwith–Wiedemann syndrome, which leads to an 800-fold increased risk of embryonal tumours — that is, those involving residual fetal tissues, such as Wilms' tumour of the kidney and rhabdomyosarcoma[50]. LOI of *IGF2* is specifically associated with increased cancer risk in children with Beckwith–Wiedemann syndrome, even though it occurs in only a fraction of the affected individuals[5] (Fig. 2). Thus, the epigenetic change precedes cancer and confers risk for cancer, a strong argument for causality. LOI of *IGF2* was found in adults, at a frequency of 5–10%[51], and is associated with a fivefold increased frequency of benign and malignant colorectal neoplasms (and 20-fold for cancer), as well as an increased family history of cancer, consistent with a causal role in cancer predisposition[52,53]. Another example of epigenetic alterations in normal tissue is the hypermethylation of *p16* that occurs with ageing[54] and in the normal tissue of women with breast cancer[55], although neither case has yet been linked to cancer risk.

Experimental data in mice further support a causal role for epigenetic changes in cancer. When *DNMT1* hypomorphs are crossed with Min (multiple intestinal neoplasia) mice with an *Apc* mutation, they show an increased frequency of intestinal neoplasia and liver cancers[56]. Hypomethylation also causes increased chromosomal instability, leading to aggressive T-cell lymphomas[57], as well as an increase in sarcomas in mice with *p53* and neurofibromin 1 (*NF1*) mutations[58]. DNA hypermethylation is also important, as *DNMT1* hypomorphs also show delayed progression of adenomas in Min mice[56,59], suggesting that hypomethylation is more important in the earliest stages of carcinogenesis, whereas hypermethylation has a greater role during tumour progression. In addition, engineered loss of one allele of *HIC1* leads to an increased number of late-onset tumours with epigenetic silencing of the remaining allele[60]. Genetically induced LOI of *IGF2* increases the frequency of adenomas in mice caused by mutations in *Apc*[61]. Moreover, engineered global LOI leads to intestinal and hepatic tumours in chimaeric mice[62].

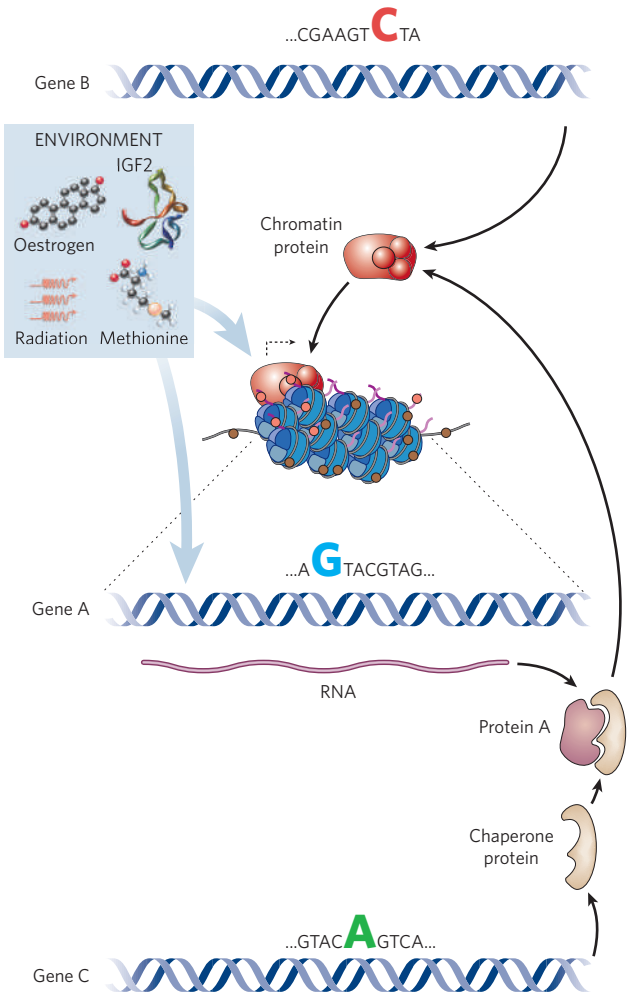


Figure 4 | The epigenome at the intersection between environment and genetic variation. According to the common disease genetic and epigenetic (CDGE) hypothesis, the epigenome may modulate the effect of genetic variation (example shown is the large nucleotide in gene A, which could be C or G), either by affecting the gene's expression through the action of chromatin proteins or DNA methylation, or by modulating protein folding of the gene product of the variant locus or chromatin protein. The epigenome may, in turn, be affected by sequence variation in the genes encoding chromatin or chaperone proteins (genes B and C, respectively). Environmental factors (such as toxins, growth factors, dietary methyl donors and hormones) can affect the genome and the epigenome.

Cancer epigenetics and the stem-cell hypothesis

Although epigenetic alterations are commonly looked on as surrogates for genetic change in cancer, they are probably also critical first steps in neoplastic progression, disrupting the normal stem- or progenitor-cell programme, for example by stimulating stem-cell proliferation outside their normal microenvironment[63]. This 'epigenetic progenitor model', in which cancer originates in stem or progenitor cells after epigenetic alterations, is supported by the ubiquitous early nature of epigenetic changes in cancer, discussed above, as well as the demonstration of altered progenitor cells in normal tissues of patients with cancer.

LOI of *IGF2* leads to an expanded progenitor-cell compartment in the intestine of mice harbouring an *Apc* mutation, and increased expression of progenitor-cell markers[61,64], a feature also seen in humans with LOI of *IGF2* and increased risk of colon cancer[61]. Similarly, LOI of *IGF2* in Beckwith–Wiedemann syndrome is specifically associated with cancer risk and leads to the expansion of nephrogenic progenitor cells[65]. Further support for the model comes from the fact that mouse melanoma and medulloblastoma nuclei can be cloned to form blastocysts or chimaeric mice[66]. Although mice derived from the former show an increased propensity to form melanomas, many of the tumour properties must be epigenetic in origin and some cells within the tumour are pluripotent[66].

In breast cancer, widespread epigenetic alterations are found in tumour cells, stromal cells and the myoepithelium, suggesting that the entire tumour microenvironment, including apparently normal cells, is the target of epigenetic disruption[67]. Cancers also seem to show increased epigenetic plasticity. This epigenetic plasticity may be an inherent property of the stem cells from which cancers arise, for example the bivalent nature of adjacent H3K4 and H3K27 methylation that is seen at many genomic sites in stem cells but not in somatic cells after differentiation[68]. The gene *MLL1*, which is rearranged and activated in childhood leukaemias, is also a key regulator of normal stem-cell differentiation[69,70]. Polycomb group genes might also be tumour progenitors, as they are overexpressed in cancer, as noted earlier, and they repress developmental regulators in embryonic stem cells[71]. Thus, epigenetics seems to be central to plasticity both in development and in tumour cells, and epigenetic discovery will be critical to understanding these.

Epigenetics and common complex disease

The next great frontier in the epigenetics of human disease is to establish its potential role in common non-neoplastic human diseases. At the moment, the most appealing candidates are disorders affecting behaviour, on the basis of clues from Rett syndrome and Prader–Willi syndrome, as well as the intriguing story of Turner syndrome in girls with only one X chromosome. Girls lacking the paternal X chromosome exhibit behavioural socialization problems more frequently than girls lacking the maternal X chromosome[72], and a candidate imprinted gene region that could be responsible has been identified[73]. Autism and bipolar disorder are two common complex traits that have defied gene identification, and both show surprisingly high frequencies of phenotypic discordance in monozygotic twins. Only 60% concordance was reported in autism using strict criteria[74], and neuroanatomical differences have been found in cerebellar grey- and white-matter volumes between discordant monozygotic twins[75]. In bipolar disorder, 30% of monozygotic twins are discordant[76], and the disease itself is episodic, with patients being seriously ill at some times and perfectly normal at others, often for long stretches of time, for no apparent reason. Some studies of both autism and bipolar disorder have also shown parent-of-origin-specific linkage, with excess transmission of paternal alleles to autism cases[77], and excess transmission of maternal alleles to bipolar disorder cases[78].

A third candidate common disease with an epigenetic component is systemic autoimmune disease. Aberrant hypomethylation is found in T cells of patients with systemic lupus erythematosis, including in genes such as lymphocyte function-associated antigen-1, which is overexpressed in lupus T cells[79]. Treatment of viable T cells with 5-aza-2′-deoxycytidine induces a syndrome in mice similar to systemic lupus erythematosis[80]. Procainamide and hydralazine both cause hypomethylation and can cause lupus, and treatment of T cells with these drugs elicits a lupus-like syndrome in mice[81].

Epigenetics and the environment

The epigenome is an important target of environmental modification. Environmental toxins such as heavy metals disrupt DNA methylation and chromatin[82]. Oestrogenic and anti-androgenic toxins that decrease male fertility alter DNA methylation, and these changes are inherited by subsequent generations[83]. Dietary modification also can have a profound effect on DNA methylation and genomic imprinting. Deficiency in folate and methionine, necessary for normal biosynthesis of S-adenosylmethionine, the methyl donor for methylcytosine, leads to aberrant imprinting of IGF2 (ref. 84), and methylation supplementation can cause methylation and silencing of a retroposon in mice with silencing of the nearby agouti coat-colour gene[85]. Colorectal cancer risk is linked to both dietary folate deficiency and variants in methylenetetrahydrofolate reductase, which has a critical role in directing the folate pool toward remethylation of homocysteine to methionine[86]. A remarkable example of an environmental effect on the epigenome is the modification of glucocorticoid receptor methylation seen in the hippocampus of rat pups in response to maternal grooming[87]. A surprising environmental modulator of the epigenome is assisted reproductive technology (ART), which has been shown to be the method of conception at higher than expected frequency in Beckwith–Wiedemann syndrome and Angelman syndrome[88]. Intriguingly, all but 1 of the 14 reported cases of Beckwith–Wiedemann syndrome associated with ART involved hypomethylation of *LIT1* (ref. 89), although this abnormality is normally present in only about one-third of patients with Beckwith–Wiedemann syndrome.

The common disease genetic and epigenetic (CDGE) hypothesis argues that in addition to genetic variation, epigenetics provides an added layer of variation that might mediate the relationship between genotype and internal and external environmental factors[90]. This epigenetic component could help to explain the marked increase in common diseases with age, as well as the frequent discordance of diseases such as bipolar disorder between monozygotic twins[76]. A common characteristic of ageing is a time-dependent decline in responsiveness or adaptation to the environment, a form of loss of phenotypic plasticity. This loss of phenotypic plasticity could be mediated epigenetically if loss of the normal balance between gene-promoting and gene-silencing factors occurred across the genome (Fig. 3). This idea is supported by a study showing greater variance of total DNA methylation and histone H3K9 acetylation in older monozygotic twins than in younger twins, although that study did not measure epigenetic changes over time in the same individual[91].

The CDGE hypothesis is also supported by two compelling lines of evidence in model organisms. First, inhibition of the chaperone protein Hsp90 in *Drosophila* leads to the expression of previously latent heritable mutations within a single meiotic cell division, which then become independent of Hsp90 (ref. 92). This mutational suppression is chromatin mediated and can be reversed by mutations in several trithorax group proteins[93]. Second, a screen for genes that cooperate in disrupting *Caenorhabditis elegans* phenotypes revealed six 'hub' genes that interacted with as many as one-quarter of all genes tested. All were components of chromatin-modifying complexes[94]. Thus, common diseases may involve phenotypic variants with both genetic variation and environmentally triggered epigenetic change that modulates the effects of DNA sequence variation (Fig. 4). These epigenetic modifiers are, in turn, affected by variation in the genes that encode them, and environmental factors (hormones, growth factors, toxins and dietary methyl donors) influence both the genome and epigenome (Fig. 4). This idea can be tested by incorporating an assessment of the epigenome into population epidemiological studies (see ref. 95 for a review), rather than simply stratifying risk for environmental exposures as is done currently.

Prospects for epigenetic therapy

As epigenetic mechanisms for human disease are identified, epigenetic therapies are being developed or rediscovered. Some drugs are used specifically because of their known effects on the epigenome. For example, two classes of epigenome-modifying agent are currently in clinical trials for cancer, for example, for the treatment of myelodysplasia[96]: DNA methyltransferase inhibitors such as decitabine, and histone deacetylase inhibitors such as SAHA (suberoylanilide hydroxamic acid). SAHA is being used for cancer treatment, although its *in vivo* targets are still unknown. The overall response rate with decitabine in a phase III study showed a small but statistically significant difference for myelodysplasia (9% complete response and 8% partial response, compared with no response in controls), and half of the clinically responsive patients showed a cytogenetic response[97]. One cautionary note about the use of nonselective agents that inhibit DNA methylation is that these drugs may activate as many genes as they silence[27]. An effect opposite to that of methyltransferase and histone deacetylase inhibitors is achieved through rational drug design of histone acetyltransferase inhibitors, for example, bisubstrate analogues such as Lys-CoA, a selective *p300/CBP* inhibitor[98]. Such drugs may be useful in cancer treatment, because *p300*-negative cells undergo increased apoptosis after chemotherapy[99].

Some drugs that have an effect on the epigenome are already in widespread use, but their epigenetic effect has only recently been discovered. For example, valproic acid is used to treat various disorders, including seizures, bipolar disorder and cancer, and valproic acid was recently found to be a potent histone deacetylase inhibitor[100]. A relatively simple drug strategy could be to target rationally designed small compounds to a epigenetically altered pathway, rather than attempting to repair an epigenetic lesion. For example, in patients with LOI of *IGF2*, the IGF2 signalling receptor, IGF1R tyrosine kinase, or the downstream Akt or ERK signalling pathways could be targeted with existing drugs or compounds under development rather than attempting to reverse the epigenetic lesion itself.

Given that epigenetics is at the heart of phenotypic variation in health and disease, it seems likely that understanding and manipulating the epigenome holds enormous promise for preventing and treating common human illness. Epigenetics also offers an important window to understanding the role of the environment's interactions with the genome in causing disease, and in modulating those interactions to improve human health. ■

1. Van Speybroeck, L. From epigenesis to epigenetics: the case of C. H. Waddington. *Ann. NY Acad. Sci.* **981,** 61–81 (2002).
2. Debaun, M. R. & Feinberg, A. P. in *Inborn Errors of Development: The Molecular Basis of Clinical Disorders of Morphogenesis* (ed. Epstein, C. J.) 758–765 (Oxford Univ. Press, Oxford, USA, 2004).
3. Niemitz, E. L. *et al.* Microdeletion of LIT1 in familial Beckwith–Wiedemann syndrome. *Am. J. Hum. Genet.* **75,** 844–849 (2004).
4. Sparago, A. *et al.* Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and Beckwith–Wiedemann syndrome. *Nature Genet.* **36,** 958–960 (2004).
5. DeBaun, M. R. *et al.* Epigenetic alterations of H19 and LIT1 distinguish patients with Beckwith–Wiedemann syndrome with cancer and birth defects. *Am. J. Hum. Genet.* **70,** 604–611 (2002).
6. Diaz-Meyer, N., Yang, Y., Sait, S. N., Maher, E. R. & Higgins, M. J. Alternative mechanisms associated with silencing of CDKN1C in Beckwith–Wiedemann syndrome. *J. Med. Genet.* **42,** 648–655 (2005).
7. Horsthemke, B. & Buiting, K. Imprinting defects on human chromosome 15. *Cytogenet. Genome Res.* **113,** 292–299 (2006).
8. Lalande, M. Imprints of disease at *GNAS1*. *J. Clin. Invest.* **107,** 793–794 (2001).
9. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genet.* **23,** 185–188 (1999).
10. Bienvenu, T. & Chelly, J. Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nature Rev. Genet.* **7,** 415–426 (2006).
11. Xu, G. L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402,** 187–191 (1999).
12. Blanco-Betancourt, C. E. *et al.* Defective B-cell-negative selection and terminal differentiation in the ICF syndrome. *Blood* **103,** 2683–2690 (2004).
13. Gibbons, R. J. & Higgs, D. R. Molecular-clinical spectrum of the ATR-X syndrome. *Am. J. Med. Genet.* **97,** 204–212 (2000).
14. Petrif, F. *et al.* Rubinstein–Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376,** 348–351 (2002).
15. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301,** 89–92 (1983).
16. Wilson, A. S., Power, B. E. & Molloy, P. L. DNA hypomethylation and human diseases. *Biochim. Biophys. Acta* **1775,** 138–162 (2007).
17. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nature Rev. Cancer* **4,** 143–153 (2004).
18. Brueckner, B. *et al.* The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res.* **67,** 1419–1423 (2007).
19. Wu, H. *et al.* Hypomethylation-linked activation of PAX2 mediates tamoxifen-stimulated endometrial carcinogenesis. *Nature* **438,** 981–987 (2005).
20. Greger, V., Passarge, E., Hopping, W., Messmer, E. & Horsthemke, B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Human Genet.* **83,** 155–158 (1989).
21. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nature Rev. Genet.* **3,** 415–428 (2002).
22. Costello, J. F. *et al.* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nature Genet.* **24,** 132–138 (2000).
23. Esteller, M., Corn, P. G., Baylin, S. B. & Herman, J. G. A gene hypermethylation profile of human cancer. *Cancer Res.* **61,** 3225–3229 (2001).
24. Frigola, J. *et al.* Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nature Genet.* **38,** 540–549 (2006).
25. Hattori, N. *et al.* Preference of DNA methyltransferases for CpG islands in mouse embryonic stem cells. *Genome Res.* **14,** 1733–1740 (2004).
26. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37,** 853–862 (2005).
27. Gius, D. *et al.* Distinct effects on gene expression of chemical and genetic manipulation of the cancer epigenome revealed by a multimodality approach. *Cancer Cell* **6,** 361–371 (2004).
28. Scrable, H. *et al.* A model for embryonal rhabdomyosarcoma tumorigenesis that involves genome imprinting. *Proc. Natl Acad. Sci. USA* **86,** 7480–7484 (1989).
29. Rainier, S. *et al.* Relaxation of imprinted genes in human cancer. *Nature* **362,** 747–749 (1993).
30. Ogawa, O. *et al.* Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* **362,** 749–751 (1993).
31. Feinberg, A. P. Genomic imprinting and gene activation in cancer. *Nature Genet.* **4,** 110–113 (1993).
32. Kondo, M. *et al.* Frequent loss of imprinting of the *H19* gene is often associated with its overexpression in human lung cancers. *Oncogene* **10,** 1193–1198 (1995).
33. van Roozendaal, C. E. *et al.* Loss of imprinting of IGF2 and not H19 in breast cancer, adjacent normal tissue and derived fibroblast cultures. *FEBS Lett.* **437,** 107–111 (1998).
34. Murphy, S. K. *et al.* Frequent *IGF2/H19* domain epigenetic alterations and elevated *IGF2* expression in epithelial ovarian cancer. *Mol. Cancer Res.* **4,** 283–292 (2006).
35. Uyeno, S. *et al. IGF2* but not *H19* shows loss of imprinting in human glioma. *Cancer Res.* **56,** 5356–5359 (1996).
36. Yuan, J. *et al.* Aberrant methylation and silencing of *ARHI*, an imprinted tumor suppressor gene in which the function is lost in breast cancers. *Cancer Res* **63,** 4174–4180 (2003).
37. Astuti, D. *et al.* Epigenetic alteration at the *DLK1–GTL2* imprinted domain in human neoplasia: analysis of neuroblastoma, phaeochromocytoma and Wilms' tumour. *Br. J. Cancer* **92,** 1574–1580 (2005).
38. Pedersen, I. S. *et al.* Frequent loss of imprinting of *PEG1/MEST* in invasive breast cancer. *Cancer Res* **59,** 5449–5451 (1999).
39. Varambally, S., *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419,** 624–629 (2002).
40. Fraga, M. F. *et al.* Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature Genet.* **37,** 391–400 (2005).
41. Tamaru, H. & Selker, E. U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414,** 277–283 (2001).
42. Terranova, R., Agherbi, H., Boned, A., Meresse, S. & Djabali, M. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc. Natl Acad. Sci. USA* **103,** 6629–6634 (2006).
43. Esteve, P. O. *et al.* Direct interaction between DNMT1and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* **20,** 3089–3103 (2006).
44. Rozenblatt-Rosen, O. *et al.* The C-terminal SET domains of ALL-1 and TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. *Proc. Natl Acad. Sci. USA* **95,** 4152–4157 (1998).
45. Versteege, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394,** 203–206 (1998).
46. Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nature Genet.* **30,** 365–366 (2002).
47. Scott, M. R., Westphal, K. H. & Rigby, P. W. Activation of mouse genes in transformed cells. *Cell* **34,** 557–567 (1983).
48. Veigl, M. L. *et al.* Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc. Natl Acad. Sci. USA* **95,** 8698–8702 (1998).
49. Bachman, K. E. *et al.* Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. *Cancer Cell* **3,** 89–95 (2003).
50. DeBaun, M. R. & Tucker, M. A. Risk of cancer during the first four years of life in children from the Beckwith–Wiedemann syndrome registry. *J. Pediatr.* **132,** 398–400 (1998).
51. Cui, H., Horon, I. L., Ohlsson, R., Hamilton, S. R. & Feinberg, A. P. Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability. *Nature Med.* **4,** 1276–1280 (1998).
52. Cui, H. *et al.* Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* **299,** 1753–1755 (2003).
53. Woodson, K. *et al.* Loss of insulin-like growth factor-II imprinting and the presence of screen-detected colorectal adenomas in women. *J. Natl Cancer Inst.* **96,** 407–410 (2004).
54. Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc. Natl Acad. Sci. USA* **96,** 8681–8686 (1999).
55. Holst, C. R. *et al.* Methylation of $p16^{INK4a}$ promoters occurs *in vivo* in histologically normal human mammary epithelia. *Cancer Res.* **63,** 1596–1601 (2003).
56. Yamada, Y. *et al.* Opposing effects of DNA hypomethylation on intestinal and liver carcinogenesis. *Proc. Natl Acad. Sci. USA* **102,** 13580–13585 (2005).
57. Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300,** 489–492 (2003).
58. Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300,** 455 (2003).
59. Laird, P. W. *et al.* Suppression of intestinal neoplasia by DNA hypomethylation. *Cell* **81,** 197–205 (1995).
60. Chen, W. Y. *et al.* Heterozygous disruption of Hic1 predisposes mice to a gender-dependent spectrum of malignant tumors. *Nature Genet.* **33,** 197–202 (2003).
61. Sakatani, T. *et al.* Loss of imprinting of *Igf2* alters intestinal maturation and tumorigenesis in mice. *Science* **307,** 1976–1978 (2005).
62. Holm, T. M. *et al.* Global loss of imprinting leads to widespread tumorigenesis in adult mice. *Cancer Cell* **8,** 275–285 (2005).
63. Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nature Rev. Genet.* **7,** 21–33 (2006).
64. Harper, J. *et al.* Soluble IGF2 receptor rescues $Apc^{Min/+}$ intestinal adenoma progression induced by *Igf2* loss of imprinting. *Cancer Res.* **66,** 1940–1948 (2006).
65. Ravenel, J.D. *et al.* Loss of imprinting of insulin-like growth factor-II (IGF2) gene in distinguishing specific biologic subtypes of Wilms tumor. *J. Natl Cancer Inst.* **93,** 1698–1703 (2001).
66. Hochedlinger, K. *et al.* Reprogramming of a melanoma genome by nuclear transplantation. *Genes Dev.* **18,** 1875–1885 (2004).
67. Hu, M. *et al.* Distinct epigenetic changes in the stromal cells of breast cancers. *Nature Genet.* **37,** 899–905 (2005).
68. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
69. Horton, S. J. *et al.* Continuous MLL-ENL expression is necessary to establish a 'Hox Code' and maintain immortalization of hematopoietic progenitor cells. *Cancer Res.* **65,** 9245–9252 (2005).
70. Krivtsov, A. V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442,** 818–822 (2006).
71. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
72. Skuse, D. H. *et al.* Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature* **387,** 705–708 (1997).
73. Raefski, A. S. & O'Neill, M. J. Identification of a cluster of X-linked imprinted genes in mice. *Nature Genet.* **37,** 620–624 (2005).
74. Bailey, A. *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.* **25,** 63–77 (1995).
75. Kates, W. R. *et al.* Neuroanatomic variation in monozygotic twin pairs discordant for the narrow phenotype for autism. *Am. J. Psychiatry* **161,** 539–546 (2004).
76. Kato, T., Iwamoto, K., Kakiuchi, C., Kuratomi, G. & Okazaki, Y. Genetic or epigenetic difference causing discordance between monozygotic twins as a clue to molecular basis of mental disorders. *Mol. Psychiatry* **10,** 622–630 (2005).
77. International Molecular Genetic Study of Autism ConsortiumConsortium. Further characterization of the autism susceptibility locus *AUTS1* on chromosome 7q. *Hum. Mol. Genet.* **10,** 973–982 (2001).
78. McInnis, M. G. *et al.* Genome-wide scan of bipolar disorder in 65 pedigrees: supportive evidence for linkage at 8q24, 18q22, 4q32, 2p12, and 13q12. *Mol. Psychiatry* **8,** 288–298 (2003).
79. Lu, Q. *et al.* Epigenetics, disease, and therapeutic interventions. *Ageing Res. Rev.* **5,** 449–467 (2006).
80. Quddus, J. *et al.* Treating activated CD4+ T cells with either of two distinct DNA methyltransferase inhibitors, 5-azacytidine or procainamide, is sufficient to cause a lupus-like disease in syngeneic mice. *J. Clin. Invest.* **92,** 38–53 (1993).

81. Richardson, B. DNA methylation and autoimmune disease. *Clin. Immunol.* **109,** 72–79 (2003).
82. Sutherland, J. E. & Costa, M. Epigenetics and the environment. *Ann. NY Acad. Sci.* **983,** 151–160 (2003).
83. Anway, M. D., Cupp, A. S., Uzumcu, M. & Skinner, M. K. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308,** 1466–1469 (2005).
84. Waterland, R. A., Lin, J. R., Smith, C. A. & Jirtle, R. L. Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (*Igf2*) locus. *Hum. Mol. Genet.* **15,** 705–716 (2006).
85. Waterland, R. A. & Jirtle, R. L. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* **23,** 5293–5300 (2003).
86. Giovannucci, E. Alcohol, one-carbon metabolism, and colorectal cancer: recent insights from molecular studies. *J. Nutr.* **134,** 2475S–2481S (2004).
87. Weaver, I. C. *et al.* Epigenetic programming by maternal behavior. *Nature Neurosci.* **7,** 847–854 (2004).
88. DeBaun, M. R., Niemitz, E. L. & Feinberg, A. P. Association of *in vitro* fertilization with Beckwith–Wiedemann syndrome and epigenetic alterations of *LIT1* and *H19*. *Am. J. Hum. Genet.* **72,** 156–160 (2003).
89. Niemitz, E. L. & Feinberg, A. P. Epigenetics and assisted reproductive technology: a call for investigation. *Am. J. Hum. Genet.* **74,** 599–609 (2004).
90. Bjornsson, H. T., Fallin, M. D. & Feinberg, A. P. An integrated epigenetic and genetic approach to common human disease. *Trends Genet.* **20,** 350–358 (2004).
91. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* **102,** 10604–10609 (2005).
92. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396,** 336–342 (1998).
93. Sollars, V. *et al.* Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nature Genet.* **33,** 70–74 (2003).
94. Lehner, B., Crombie, C., Tischler, J., Fortunato, A. & Fraser, A. G. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genet.* **38,** 896–903 (2006).
95. Callinan, P. A. & Feinberg, A. P. The emerging science of epigenomics. *Hum. Mol. Genet.* **15,** R95–R101 (2006).
96. Mack, G. S. Epigenetic cancer therapy makes headway. *J. Natl Cancer Inst.* **98,** 1443–1444 (2006).
97. Yee, K. W., Jabbour, E., Kantarjian, H. M. & Giles, F. J. Clinical experience with decitabine in North American patients with myelodysplastic syndrome. *Ann Hematol* **84** (Suppl. 13)**,** 18–24 (2005).
98. Zheng, Y. *et al.* Selective HAT inhibitors as mechanistic tools for protein acetylation. *Methods Enzymol.* **376,** 188–199 (2004).
99. Iyer, N. G., Ozdag, H. & Caldas, C. p300/CBP and cancer. *Oncogene* **23,** 4225–4231 (2004).
100. Phiel, C. J. *et al.* Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. *J. Biol. Chem.* **276,** 36734–36741 (2001).

Acknowledgements I thank H. Bjornsson, R. Ohlsson, T. Ekstrom, D. Gius and C. Ladd-Acosta for their many thoughtful insights, and J. Fairman for her artistry. This work was supported by a grant from the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (afeinberg@jhu.edu).

ARTICLES

Single-exciton optical gain in semiconductor nanocrystals

Victor I. Klimov[1], Sergei A. Ivanov[1], Jagjit Nanda[1], Marc Achermann[1], Ilya Bezel[1], John A. McGuire[1] & Andrei Piryatinski[1]

Nanocrystal quantum dots have favourable light-emitting properties. They show photoluminescence with high quantum yields, and their emission colours depend on the nanocrystal size—owing to the quantum-confinement effect—and are therefore tunable. However, nanocrystals are difficult to use in optical amplification and lasing. Because of an almost exact balance between absorption and stimulated emission in nanoparticles excited with single electron–hole pairs (excitons), optical gain can only occur in nanocrystals that contain at least two excitons. A complication associated with this multiexcitonic nature of light amplification is fast optical-gain decay induced by non-radiative Auger recombination, a process in which one exciton recombines by transferring its energy to another. Here we demonstrate a practical approach for obtaining optical gain in the single-exciton regime that eliminates the problem of Auger decay. Specifically, we develop core/shell hetero-nanocrystals engineered in such a way as to spatially separate electrons and holes between the core and the shell (type-II heterostructures). The resulting imbalance between negative and positive charges produces a strong local electric field, which induces a giant (~100 meV or greater) transient Stark shift of the absorption spectrum with respect to the luminescence line of singly excited nanocrystals. This effect breaks the exact balance between absorption and stimulated emission, and allows us to demonstrate optical amplification due to single excitons.

Numerous technologies—including optical interconnects in microelectronics, lab-on-a-chip chemo- and bio-analyses, optical telecommunications and information processing—would greatly benefit from flexible, chemically processable optical-gain materials that could be manipulated using simple solution-based techniques. One class of such materials is colloidal semiconductor nanocrystals, also known as nanocrystal quantum dots[1,2]. These are nanoscale crystalline particles surrounded by a layer of organic ligand molecules. The dual inorganic–organic nature of these structures provides great flexibility for controlling their physical and chemical properties. For example, using the quantum-confinement effect, the nanocrystal emission energy can be tuned by hundreds of milli-electron volts by simply changing the inorganic-core size[3]. On the other hand, relatively straightforward surface chemistry can be applied to tune nanocrystal chemical reactivity to facilitate their incorporation into, for example, nanophotonic or nanoplasmonic feedback structures for fabricating micro-lasers of various configurations[4,5].

Well-passivated nanocrystals are characterized by near-unity photoluminescence quantum yields. However, despite their high emission efficiencies, there are significant challenges to practical applications of nanocrystals in lasing technologies. Because of the degeneracy of the lowest-energy emitting levels, population inversion in nanocrystals can only be achieved if the average number of electron–hole pairs (excitons) per nanocrystal, $\langle N \rangle$, is greater than 1, which implies that at least some of the nanocrystals in the sample must contain multiexcitons[6]. A significant complication arising from this multiexcitonic nature of optical amplification in nanocrystals is highly efficient non-radiative Auger recombination induced by confinement-enhanced exciton–exciton (X–X) interactions. This process results in fast optical-gain decay characterized by picosecond timescales[7]. Demonstrated approaches to reducing Auger rates include the use of elongated nanocrystals (quantum rods)[8,9] or core/shell heteronanocrystals[10,11] that allow X–X coupling to be decreased without losing the benefits of strong quantum confinement. However, the most radical strategy for solving the problem of Auger decay is the development of methods and/or structures that would allow realization of optical gain in the single-exciton regime, for which Auger recombination is simply inactive.

Here we report a successful practical implementation of one such method that makes use of type-II core/shell heteronanocrystals. These heterostructures are engineered in such a way as to separately confine electrons and holes in the core and the shell, respectively. Spatial separation between negative and positive charges results in a strong local electric field, which leads to effective splitting of the degeneracy of the lowest-energy transition by the Stark effect and displaces the absorbing band in singly excited nanocrystals with respect to the emission line. By significantly reducing absorption losses, this effect allows for optical amplification in the single-exciton regime.

The concept of single-exciton gain

Optical gain corresponds to a light–matter interaction regime for which generation of photons by stimulated emission dominates over photon absorption. As in other lasing media, optical gain in nanocrystals requires population inversion—that is, the situation in which the number of electrons in the excited state is greater than that in the ground state. The lowest-energy emitting transition in nanocrystals of II–VI semiconductors studied here can be described in terms of a two-level system that has two electrons in its ground state. Excitation of a single electron (single exciton) across the energy gap (E_g) in this system does not produce optical gain but rather results in optical transparency, for which stimulated emission by a conduction-band electron is exactly compensated by absorption due to the electron remaining in the valence band (Fig. 1a). Stimulated emission

[1]Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

dominates over absorption only if the second electron is also excited across the energy gap, indicating that optical gain requires doubly excited nanocrystals, that is, biexcitons. These considerations imply that population inversion in a nanocrystal ensemble can only be achieved if $\langle N \rangle$ is greater than 1.

The condition for optical gain however changes if one accounts for a local electric field associated with an excited electron–hole pair. This field can alter the absorption energy of the electron remaining in the valence band by the carrier-induced Stark effect (Fig. 1b)[12,13]. If the magnitude of the Stark shift (Δ_S) is comparable to or greater than the transition line width (Γ), it can completely eliminate absorption losses at the emission wavelength in excited nanocrystals, which should allow optical gain using single-exciton states. Specifically, the threshold for population inversion in the presence of the transient transition shift is determined by the condition $\langle N \rangle = 2/\left(3 - \exp\left(-\Delta_S^2/\Gamma^2\right)\right)$ (see Supplementary Information). If $\Delta_S \ll \Gamma$, it reduces to $\langle N \rangle = 1$, which corresponds to the usual case of multiexciton optical gain. However, if $\Delta_S \gg \Gamma$, $\langle N \rangle = 2/3$ (Fig. 1c), which implies that optical gain does not require multiexcitons.

Engineered exciton–exciton interactions

The carrier-induced Stark effect can be described in terms of the Coulomb interaction of the initially generated exciton with the exciton created in the second excitation act. In this description, the transient Stark shift is determined by the X–X Coulomb interaction energy ($\Delta_{XX} = \Delta_S$) defined as $\Delta_{XX} = E_{XX} - 2E_X$, where E_X and E_{XX} are single- and biexciton energies, respectively. The X–X interaction strength is also often characterized in terms of the biexciton binding energy (δE_{XX}), which relates to Δ_{XX} by $\delta E_{XX} = -\Delta_{XX}$.

The energy Δ_{XX} depends on the local electrical charge density $\rho_X(\mathbf{r})$ associated with a single-exciton state and, hence, on the sum of the hole (ρ_h) and the electron (ρ_e) charge densities: $\rho_X(\mathbf{r}) = \rho_h(\mathbf{r}) + \rho_e(\mathbf{r})$ ($\mathbf{r}$ is the spatial coordinate). Because of almost identical spatial distributions of electron (ψ_e) and hole (ψ_h) wavefunctions, $\rho_X(\mathbf{r})$ is nearly zero in homogeneous nanocrystals ($\rho_X(\mathbf{r}) = e\left(|\psi_h(\mathbf{r})|^2 - |\psi_e(\mathbf{r})|^2\right) \approx 0$, e is the electron charge) (Fig. 2a), which leads to relatively small X–X interaction energies of ~10 to ~30 meV (refs 14, 15). These values are smaller than typical transition line widths in existing nanocrystal samples (ensemble broadening of ~100 meV or greater) and, therefore, do not allow significant suppression of absorption at the emission wavelength.

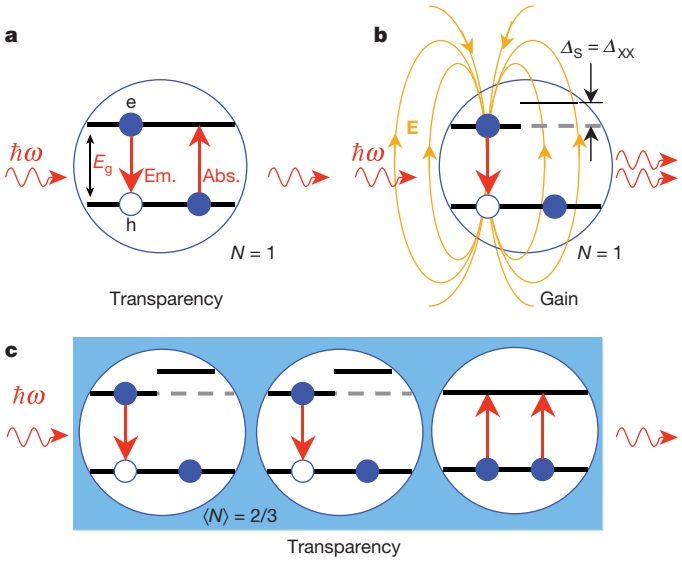


Figure 1 | The concept of single-exciton nanocrystal lasing. **a**, In the absence of X–X interactions, excitation of a single electron–hole pair (single exciton) per nanocrystal (NC) on average does not produce optical gain but results in optical transparency—that is, the regime for which stimulated emission (Em.) is exactly compensated by absorption (Abs.). e, Electron; h, hole. **b**, The balance between stimulated emission and absorption is broken if one accounts for X–X interactions that spectrally displace the absorbing transition with respect to the emission band. The latter effect can be interpreted in terms of the transition Stark shift ($\Delta_S = \Delta_{XX}$) induced by a local electric field (**E**) associated with a single-exciton state. If the transition shift is greater than the ensemble line width, optical gain can occur in the single-exciton regime. **c**, In the case of large Δ_S ($\Delta_S \gg \Gamma$), stimulated emission in singly excited NCs (n_x is their fraction in the NC ensemble) competes only with absorption in unexcited NCs (fraction $(1-n_x)$; here, we neglect multiexcitons). The stimulated-emission cross-section of a singly excited NC is one half of the absorption cross-section of the lowest-energy NC transition. Based on these considerations, the optical-gain threshold can be found from the condition $n_x/2 = (1-n_x)$, which indicates that the single-exciton gain onset corresponds to the situation where two-thirds of the NCs are excited with single excitons.

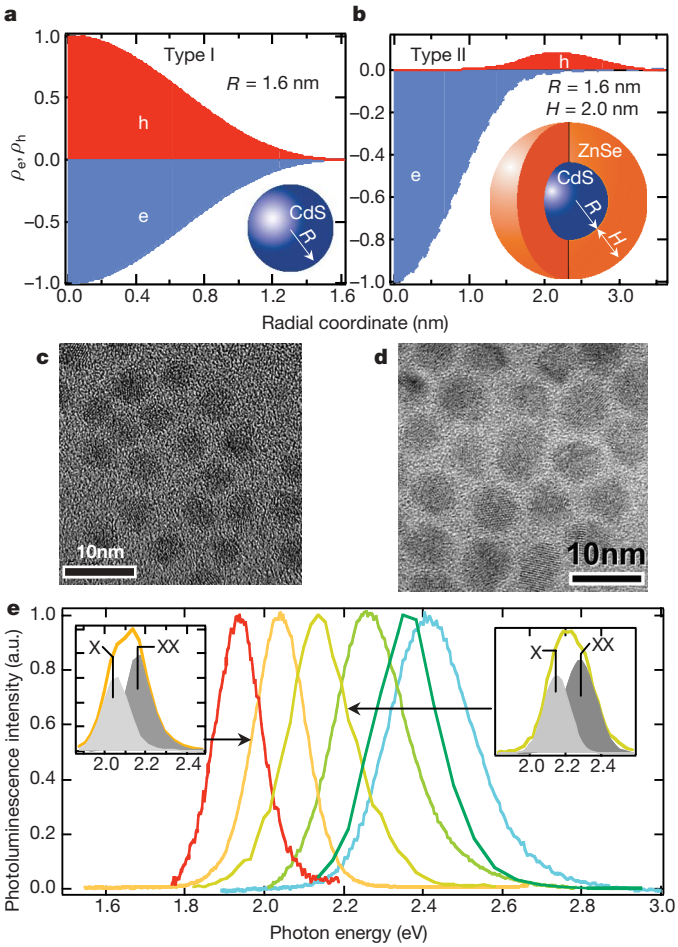


Figure 2 | Type-I CdS NCs and type-II CdS/ZnSe core/shell heteronanostructures. **a**, Spatial distributions of electron (ρ_e) and hole (ρ_h) charge densities are nearly identical in type-I NCs (inset) (calculated for $R = 1.6$ nm). **b**, In type-II core/shell NCs (inset), electrons are primarily localized in the core, while holes reside in the shell, which leads to a significant difference in radial distributions of ρ_e and ρ_h (calculated for $R = 1.6$ nm and $H = 2$ nm). **c**, A transmission electron microscopy (TEM) image of a core CdS particle of ~2.4 nm radius. **d**, TEM images of hetero-NCs fabricated using the cores shown in **c**. The core/shell NC radius is ~3.4 nm, indicating that the shell width is ~1.0 nm. Because of the small difference between electron scattering cross-sections of CdS and ZnSe, the CdS/ZnSe interface is not discernable in the TEM image. **e**, Emission spectra of a series of CdS/ZnSe core/shell NCs synthesized using CdS cores with radii of 1.6 nm (higher-energy emission) and 2.6 nm (lower-energy emission) and various shell widths. Insets, early-time emission spectra measured at high pump intensities can be deconvolved into the single-exciton (X) and the biexciton (XX) bands. The spacing between these two bands indicates giant X–X interaction energies of more than 100 meV. a.u., arbitrary units.

The separation of electrons and holes between the core and the shell in type-II nanocrystals (Fig. 2b) can lead to sizable local charge densities and, hence, large Coulomb interaction energies[16]. To analyse the effect of charge separation on X–X interactions and its influence on optical gain properties of nanocrystals, we study heteronanostructures composed of a CdS core overcoated with a ZnSe shell (Fig. 2b inset). These nanostructures are synthesized by reacting prefabricated CdS core particles[17] with Zn/Se precursors, which results in CdS(core)/ZnSe(shell) nanocrystals (Figs 2c, d) (details of the synthesis are described in Supplementary Information). As illustrated in Fig. 2e, the emission colour produced by these nanocrystals can be tuned from red to green by varying the core radius and/or the shell thickness.

According to bulk-semiconductor parameters, the bottom of the conduction band is lower in CdS than in ZnSe, while the top of the valence band is higher for ZnSe (Fig. 3a). Therefore, the electron–hole pair generated near the bulk CdS/ZnSe hetero-interface tends to produce a charge-separated state, with the electron residing in CdS and the hole in ZnSe. This situation corresponds to the type-II regime, whereas the regime for which the electrons and the holes co-occupy the same part of a heterostructure is usually referred to as type I. In contrast to a fixed alignment of energy levels at the bulk CdS/ZnSe hetero-interface, the alignment of energy states in CdS/ZnSe nanocrystals depends on the core radius (R) and the shell width (H), which determine the positions of quantized levels with respect to the bulk band edges (see the 'localization' phase diagram in Fig. 3b). Specifically, for small core radii ($R < 1.2$–1.5 nm) and thin shells ($H < 0.6$–1.0 nm), these structures yield type-I localization (shaded area in the lower-left corner of Fig. 3b), while the type of localization changes to type II for larger R and H (shaded area in the upper-right corner of Fig. 3b). For the type-I regime, an electron and a hole are delocalized over the entire heteronanocrystal and their charge densities nearly cancel each other (Fig. 2a). In the type-II case, electrons reside in the core while holes are in the shell, which leads to a significant difference in the spatial distributions of negative and positive charges and, hence, large local charge densities (Fig. 2b).

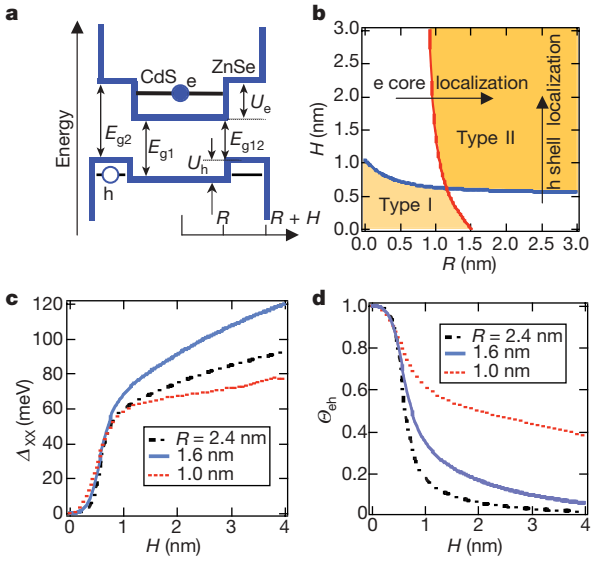


Figure 3 | Electronic structure and different localization regimes in CdS/ZnSe core/shell NCs. a, Illustration of alignment of the conduction- and valence-band edges at the bulk CdS/ZnSe interface (blue lines) and the lowest-energy electron and hole quantized levels (black lines) in the CdS(core)/ZnSe(shell) NC. Arrows indicate bulk parameters: $E_{g1} = 2.485$ eV (CdS energy gap), $E_{g2} = 2.72$ eV (ZnSe energy gap), $E_{g12} = 1.925$ eV ('indirect' energy gap of a type-II CdS/ZnSe heterostructure), $U_e = 0.795$ eV (conduction-band energy offset at the hetero-interface) and $U_h = 0.56$ eV (valence-band energy offset at the hetero-interface). **b**, Localization phase diagram that shows the regions of (R,H)-space that correspond to different localization regimes in CdS/ZnSe hetero-NCs. The shaded areas correspond to type-I (both carriers are delocalized over the entire hetero-NC volume) and type-II (electrons and holes are confined to the core and the shell, respectively) regimes. In the unshaded areas, the regime of localization is not well defined; in this case one of the carriers is confined to either the core or the shell, while the other one is delocalized over the entire NC volume. **c**, **d**, The dependence of the X–X interaction energy Δ_{XX} (in **c**) and the electron–hole overlap integral Θ_{eh} (in **d**) on shell width for three core radii of 1.0, 1.6 and 2.4 nm. For all core radii, the increase in Δ_{XX} correlates with the drop in Θ_{eh}, indicating that X–X interactions are enhanced with increasing degree of spatial separation between electrons and holes. From the data for $R = 1.0$ and 1.6 nm, one can see that the more complete spatial separation achievable for 1.6-nm cores produces stronger X–X repulsion. The trend is more complex for larger core sizes (compare data for $R = 1.6$ and 2.4 nm), for which the increase in the interaction energy due to increasing charge imbalance competes with the effect of the decreasing charge density.

A convenient quantity for describing the spatial separation between electrons and holes is the electron–hole overlap integral, $\Theta_{eh} = |\langle \psi_h \mid \psi_e \rangle|^2$. This quantity also provides a measure of the imbalance between negative and positive charges in the nanocrystal and, therefore, the changes in Θ_{eh} directly correlate with variations in the X–X interaction energy. This effect is illustrated in Fig. 3, which shows Δ_{XX} (Fig. 3c) and Θ_{eh} (Fig. 3d) calculated for CdS/ZnSe nanocrystals for fixed core radii ($R = 1.0$, 1.6 and 2.4 nm) and a varied shell width (details of the calculations are described in Supplementary Information). We observe that the initial increase in H leads to a rapid drop of the overlap integral owing to transition to the type-II regime. The reduction in Θ_{eh} is accompanied by a rapid increase of Δ_{XX} because of an increasing imbalance between positive and negative charges. Specifically, for $R = 1.6$ nm and $H > 2$ nm ($\Theta_{eh} < 0.17$), Δ_{XX} can reach giant values of $\sim$100 meV.

It is interesting to examine the sign of the X–X interaction energy. In type-I nanocrystals, Coulomb interactions tend to spatially arrange charges in such a way that the biexciton energy E_{XX} is less than twice the single-exciton energy E_X. This situation corresponds to a negative value of Δ_{XX} (positive biexciton binding energy), which can be interpreted in terms of an effective X–X attraction. In type-II nanocrystals, the spatial distribution of charges is controlled not by Coulomb interactions but by large energy gradients at the core/shell interface, which leads to concentration of the same-sign charges in the same part of the heteronanocrystal (both electrons in the core; both holes in the shell) and spatial separation of charges of the opposite sign across the hetero-interface. This type of spatial arrangement increases the repulsive component of the Coulomb interaction and decreases its attractive component, which produces net X–X repulsion[16] (negative biexciton binding energy), as indicated by the positive sign of the calculated values of Δ_{XX} (Fig. 3c). X–X repulsion can also be obtained in type-I nanocrystals if one accounts for the difference between the conduction- and the valence-band structures[18]. However, the latter repulsion is significantly smaller than the repulsion discussed in this Article, and so is neglected in our calculations.

The sign of Δ_{XX} has an important effect on the optical-gain properties of nanocrystals because it determines the direction of the shift of the absorbing transitions with respect to the emission line. If Δ_{XX} is negative (X–X attraction), the transitions move downward in energy, which may have a detrimental effect on lasing performance because of increasing absorption due to the manifold of strong transitions located immediately above the emitting band. On the other hand, strong X–X repulsion, which can be produced in type-II nanocrystals, should benefit lasing because it moves strongly absorbing transitions away from the emission line.

Giant exciton–exciton repulsion energies

In order to experimentally measure the X–X interaction energy in type-II core/shell nanocrystals, we compare the position of the biexciton photoluminescence band with respect to the single-exciton line. Radiative recombination of the biexciton produces a photon

($\hbar\omega_{XX}$) and an exciton and hence, $\hbar\omega_{XX} = E_{XX} - E_X = E_X + \Delta_{XX}$. On the basis of this expression, the shift of the biexciton line with respect to the single-exciton band ($\hbar\omega_X = E_X$) provides a direct measure of the X–X interaction energy: $\Delta_{XX} = \hbar\omega_{XX} - \hbar\omega_X$.

The challenge in experimentally detecting photoluminescence from nanocrystal multiexcitons is associated with their short (picoseconds to hundreds of picoseconds) lifetimes, which are limited by non-radiative Auger recombination[7]. Because these times are significantly shorter than the radiative time constants, signals from multiexcitons are not well pronounced in steady-state photoluminescence spectra. Therefore, in order to detect the emission from multiexcitons, we apply time-resolved photoluminescence measurements, in which emission of nanocrystals excited by 200-fs, 3-eV pulses is analysed using time-correlated single-photon counting (~30 ps time resolution).

Figure 4a shows photoluminescence spectra of a hexane solution of CdS/ZnSe nanocrystals with $R = 1.6$ nm and $H = 2$ nm recorded at a pump fluence that corresponds to excitation of approximately 1.5 excitons per nanocrystal on average. According to the phase diagram in Fig. 3b, the nanocrystals used in these measurements correspond to the type-II regime, for which electrons and holes are well separated between the core and the shell. The long-time spectrum recorded at time $t = 10$ ns after excitation (filled circles in Fig. 4a) is identical to the steady-state photoluminescence spectrum observed at low excitation fluences ($\langle N \rangle \ll 1$) and corresponds to emission of single excitons ($E_X = 2.054$ eV). The $t = 0$ photoluminescence (open squares in Fig. 4a) indicates the presence of an additional high-energy, short-lived band at 2.160 eV, which decays with a time constant of ~130 ps. This band (feature XX in Fig. 4a) can be extracted from the $t = 0$ spectrum by subtracting an appropriately scaled long-time spectrum (feature X in Fig. 4a).

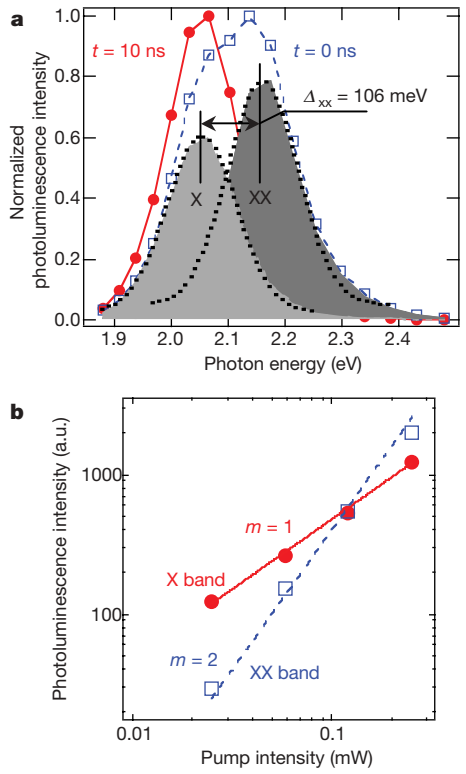


Figure 4 | Experimental demonstration of giant X–X repulsion in type-II CdS/ZnSe NCs. **a**, Short- (open squares) and long- (solid circles) time room-temperature photoluminescence spectra of CdS/ZnSe NCs. Deconvolution of the $t = 0$ spectrum into two bands shown by grey areas (dotted lines are gaussian fits) reveals the short-lived XX feature due to emission from biexcitons that is located to high energies from the single-exciton X band. The spacing between these two bands indicates giant X–X repulsion characterized by $\Delta_{XX} = 106$ meV. **b**, Photoluminescence pump-intensity dependences indicate that the growth of X and XX features with excitation power shows log–log slopes of $m = 1$ and 2, respectively. These slopes are consistent with single-exciton and the biexciton mechanisms for the X and XX emission bands, respectively.

The measured pump-intensity dependences indicate that the growth of band X is linear in pump fluence (filled circles in Fig. 4b), as expected for single-exciton emission. On the other hand, the high-energy band XX shows a quadratic growth (open squares in Fig. 4b), which is typical for emission from biexcitons. The fast decay of this feature is also consistent with its multiexciton origin, and is due to efficient Auger recombination. Further, this band cannot be attributed to recombination of carriers populating excited nanocrystal states because, according to our transient absorption results (not shown), the first optical transition involving excited electronic states in these nanocrystals is located at ~2.254 eV, which is nearly 100 meV higher than the XX feature. On the basis of these considerations, the band XX can be assigned to emission from a biexciton, which comprises two lowest-energy excitons.

The fact that the biexciton emission occurs at higher energies than the single-exciton photoluminescence indicates the repulsive character of the X–X interaction, as predicted by our modelling. Further, on the basis of the measured spectral positions of bands X and XX, we obtain $\Delta_{XX} = 106$ meV, which is in good agreement with the value of 91 meV calculated for nanocrystals with the geometrical parameters studied here. Strong X–X repulsion ($\Delta_{XX} > 100$ meV) is consistently observed for type-II CdS/ZnSe nanocrystal samples with different emission colours, as illustrated in the two insets of Fig. 2e. We also detect large X–X interaction energies (~80 meV) for type-II nanocrystals of a different composition (ZnTe(core)/CdSe(shell)) with emission in the near-infrared (770 nm). All of these observations point towards the generality of giant X–X repulsion energies in strongly confined type-II colloidal nanoparticles.

Single-exciton optical gain and ASE

To analyse light amplification in type-II CdS/ZnSe nanocrystals, we compare their optical-gain properties with those of traditional type-I CdSe nanocrystals with matching emission wavelengths (Fig. 5). For the biexcitonic gain mechanism operating in type-I nanocrystals, a sharp peak of amplified spontaneous emission (ASE) is red-shifted with respect to the single-exciton band[6,19] (Fig. 5a, upper spectrum) because of X–X attraction, which decreases the emission energy of biexcitons compared to that of single excitons. For type-II samples, we observe that as we increase the pump level a new, sharp emission feature develops near the position of the single-exciton band (2.01 eV) (Fig. 5b, and the lower spectrum in Fig. 5a). This new peak shows a clear excitation threshold of ~2 mJ cm^{-2} and a fast, superlinear growth with increasing pump fluence (Fig. 5b, c). The development of a similar sharp feature is also detected using a fixed pump fluence and increasing the size of the excitation spot (variable-stripe-length configuration[20]). These behaviours are consistent with the ASE regime. An important observation is that the ASE peak develops near the centre of band X, indicating that it is due to stimulated emission of single excitons. This assignment is further confirmed by the observation of a second ASE feature at higher fluences (excitation threshold of ~6 mJ cm^{-2}), which develops near the position of the high-energy XX band and is due to the traditional biexcitonic gain mechanism (Fig. 5b and c). Similar trends are observed for type-II nanocrystals with other emission wavelengths. For example, for a sample emitting at ~2.21 eV, we first detect the emergence of a single-exciton ASE feature, which is followed by the development of a bluer biexcitonic ASE band at ~2.36 eV at higher pump fluences.

To further verify the single-exciton character of light amplification in type-II nanocrystals, we perform direct measurements of optical gain using a transient absorption experiment. In these measurements, the absorption change ($\Delta\alpha$) induced in a sample by a short, 100-fs pulse is monitored using a broadband pulse of a femtosecond white-light continuum[21]. The transition to optical gain corresponds

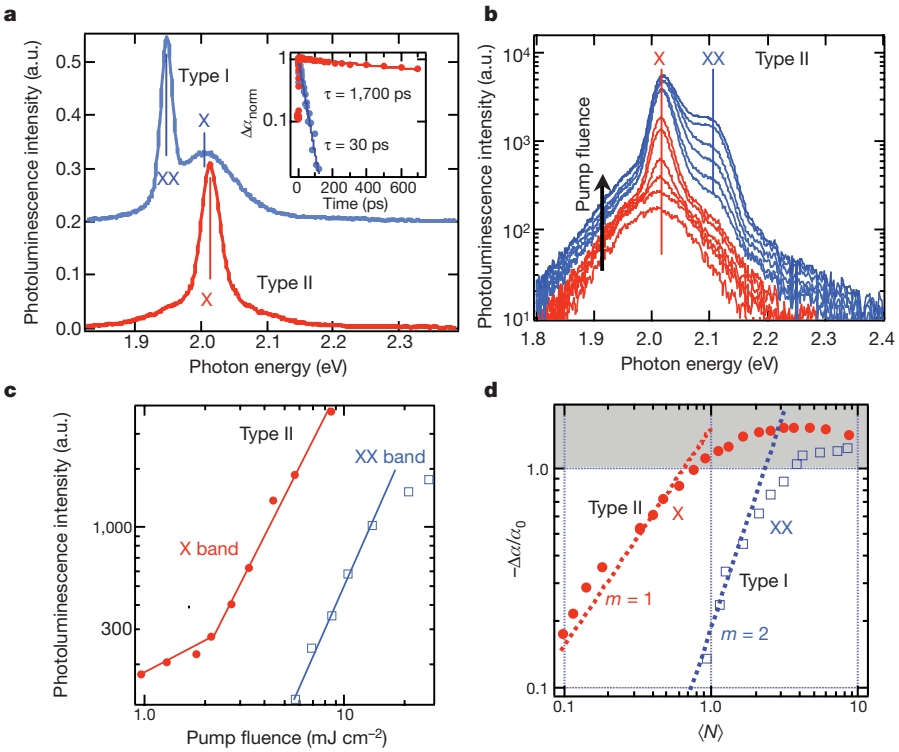


Figure 5 | Optical amplification in type-I and type-II NCs. a, Room-temperature ASE spectra of type-I CdSe NCs (blue line; offset vertically for clarity) and type-II CdS/ZnSe NCs (red line; sample is similar to the one shown in Fig. 4) prepared in the form of close-packed drop-cast films and excited by 100-fs pulses at 3 eV. Inset, transient-absorption dynamics measured for these NCs (colour-matched symbols) at the position of the ASE band for the pump intensity that corresponds to the optical-gain onset. **b**, The pump-intensity-dependent photoluminescence spectra of the type-II sample show the development of a narrow ASE peak near the centre of the single-exciton emission band. The second ASE band, which develops at higher fluences, is located near the XX photoluminescence feature. **c**, Both ASE features show superlinear pump dependence above the threshold pump intensities of ~2 mJ cm^{-2} (band X) and ~6 mJ cm^{-2} (band XX). **d**, The dependence of normalized absorption bleaching at the position of the ASE band in type-I (squares) and type-II (circles) NCs (measured ~3 ps after excitation) as a function of $\langle N \rangle$ in comparison with linear (red straight line) and quadratic (blue straight line) growth. The area shown in grey corresponds to optical gain ($-\Delta\alpha/\alpha_0 > 1$).

to the situation for which absorption bleaching ($\Delta\alpha < 0$) becomes greater than absorption of an unexcited sample (α_0). Figure 5d shows the dependence of the normalized absorption bleaching ($-\Delta\alpha/\alpha_0$) on the average number of excitons per nanocrystal for type-I (squares) and type-II (circles) nanocrystals. In addition to a lower gain threshold (factor of ~5 difference), type-II structures clearly show a different functional dependence in the development of the optical gain. In the type-II sample, the initial growth of $|\Delta\alpha|$ is nearly linear (straight red line), whereas in the type-I nanocrystals it is close to quadratic (straight blue line). These observations are consistent with single-exciton (type II) and biexciton (type I) mechanisms of optical gain in these two types of nanostructures.

The single-exciton-gain regime demonstrated here should significantly simplify real-life applications of chemically synthesized nanocrystals in lasing technologies, and specifically should allow realization of nanocrystal lasing under continuous-wave (c.w.) excitation. The pump intensity threshold for producing c.w. gain is approximately determined by the ratio of the threshold fluence measured using ultrafast excitation (see, for example, Fig. 5c) and the gain lifetime[22]. Because of Auger recombination, this lifetime is in the sub-100 ps range for the multiexcitonic gain mechanism, which leads to very high c.w. lasing thresholds that are well above the nanocrystal-photostability limit. For single-exciton gain, the intrinsic gain dynamics is determined by the radiative single-exciton lifetime, which is typically orders of magnitude longer than the Auger-decay time constants. The difference in relaxation behaviour for single- and biexciton gain mechanisms is illustrated in the inset of Fig. 5a, which shows relaxation dynamics of $\Delta\alpha$ measured at the onset of optical gain for type-I (blue circles) and type-II (red circles) samples. In type-I nanocrystals the measured decay time is 30 ps, whereas it is more than 50 times longer (1,700 ps) for the type-II nanocrystals.

The present work represents the first practical demonstration of nanocrystal structures that produce optical amplification due to stimulated emission of single-exciton states, which eliminates complications associated with ultrafast multiexciton Auger recombination. This new approach makes use of dynamic splitting of the degeneracy of the lowest-energy emitting transition by giant X–X interactions that develop in type-II heterostructures following spatial separation of electrons and holes. Implementation of the single-exciton gain regime could allow reduction of the lasing threshold under c.w. excitation by orders of magnitude, which could significantly enhance the technological potential of colloidal nanocrystals as 'soft', chemically processable optical-gain media.

Received 12 December 2006; accepted 12 April 2007.

1. Alivisatos, A. P. Semiconductor clusters, nanocrystals, and quantum dots. *Science* **271**, 933–937 (1996).
2. Klimov, V. I. (ed.) *Semiconductor and Metal Nanocrystals: Synthesis and Electronic and Optical Properties* (Marcel Dekker, New York, 2003).
3. Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E=S, Se, Te) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993).
4. Eisler, H.-J. *et al.* Color-selective semiconductor nanocrystal laser. *Appl. Phys. Lett.* **80**, 4614–4616 (2002).
5. Malko, A. V. *et al.* From amplified spontaneous emission to microring lasing using nanocrystal quantum dot solids. *Appl. Phys. Lett.* **81**, 1303–1305 (2002).
6. Klimov, V. I. *et al.* Optical gain and stimulated emission in nanocrystal quantum dots. *Science* **290**, 314–317 (2000).
7. Klimov, V. I., Mikhailovsky, A. A., McBranch, D. W., Leatherdale, C. A. & Bawendi, M. G. Quantization of multiparticle Auger rates in semiconductor quantum dots. *Science* **287**, 1011–1013 (2000).

8. Kazes, M., Lewis, D. Y., Evenstein, Y., Mokari, T. & Banin, U. Lasing from semiconductor quantum rods in a cylindrical microcavity. *Adv. Mater.* **14,** 317–321 (2002).
9. Htoon, H., Hollingsworth, J. A., Dickerson, R. & Klimov, V. I. Zero- to one-dimensional transition and Auger recombination in semiconductor quantum rods. *Phys. Rev. Lett.* **91,** 227401 (2003).
10. Ivanov, S. A. *et al.* Light amplification using inverted core/shell nanocrystals: Towards lasing in the single-exciton regime. *J. Phys. Chem. B* **108,** 10625–10630 (2004).
11. Nanda, J. *et al.* Absorption cross sections and Auger recombination lifetimes in inverted core/shell nanocrystals: Implications for lasing performance. *J. Appl. Phys.* **99,** 034309 (2006).
12. Norris, D. J., Sacra, A., Murray, C. B. & Bawendi, M. G. Measurement of the size-dependent hole spectrum in CdSe quantum dots. *Phys. Rev. Lett.* **72,** 2612–2615 (1994).
13. Klimov, V. I. Optical nonlinearities and ultrafast carrier dynamics in semiconductor nanocrystals. *J. Phys. Chem. B* **104,** 6112–6123 (2000).
14. Kang, K. I. *et al.* Confinement-enhanced biexciton binding energy in semiconductor quantum dots. *Phys. Rev. B* **48,** 15449–15452 (1993).
15. Achermann, M., Hollingsworth, J. A. & Klimov, V. I. Multiexcitons confined within a subexcitonic volume: Spectroscopic and dynamical signatures of neutral and charged biexcitons in ultrasmall semiconductor nanocrystals. *Phys Rev. B* **68,** 245302 (2003).
16. Piryatinski, A., Ivanov, S. A., Tretiak, S. & Klimov, V. I. Effect of quantum and dielectric confinement on the exciton-exciton interaction energy in type-II core/shell semiconductor nanocrystals. *Nano Lett.* **7,** 108–115 (2007).
17. Cao, Y. C. & Wang, J. H. One-pot synthesis of high-quality zinc-blende CdS nanocrystals. *J. Am. Chem. Soc.* **126,** 14336–14337 (2004).
18. Efros, A. L. & Rodina, A. V. Confined excitons, trions, and biexcitons in semiconductor microcrystals. *Solid State Commun.* **72,** 645–649 (1989).
19. Schaller, R. D., Petruska, M. A. & Klimov, V. I. Tunable near-infrared optical gain and amplified spontaneous emission using PbSe nanocrystals. *J. Phys. Chem. B* **107,** 13765–13768 (2003).
20. Shaklee, K. L., Nahory, R. E. & Leheny, R. F. Optical gain in semiconductors. *J. Lumin.* **7,** 284–309 (1973).
21. Klimov, V. I. & McBranch, D. W. Femtosecond high-sensitivity, chirp-free transient absorption spectroscopy using kilohertz lasers. *Opt. Lett.* **23,** 277–279 (1998).
22. Mikhailovsky, A. A., Malko, A. V., Hollingsworth, J. A., Bawendi, M. G. & Klimov, V. I. Multiparticle interactions and stimulated emission in chemically synthesized quantum dots. *Appl. Phys. Lett.* **80,** 2380–2382 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Chemical Sciences, Biosciences and Geosciences Division of the Office of Basic Energy Sciences, US Department of Energy (DOE), Los Alamos LDRD funds, and the Center for Integrated Nanotechnologies jointly operated for DOE by Los Alamos and Sandia National Laboratories.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.I.K. (klimov@lanl.gov).

ARTICLES

OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells

Irina V. Kovtun[1], Yuan Liu[5], Magnar Bjoras[4], Arne Klungland[4], Samuel H. Wilson[5] & Cynthia T. McMurray[1,2,3]

Although oxidative damage has long been associated with ageing and neurological disease, mechanistic connections of oxidation to these phenotypes have remained elusive. Here we show that the age-dependent somatic mutation associated with Huntington's disease occurs in the process of removing oxidized base lesions, and is remarkably dependent on a single base excision repair enzyme, 7,8-dihydro-8-oxoguanine-DNA glycosylase (OGG1). Both *in vivo* and *in vitro* results support a 'toxic oxidation' model in which OGG1 initiates an escalating oxidation–excision cycle that leads to progressive age-dependent expansion. Age-dependent CAG expansion provides a direct molecular link between oxidative damage and toxicity in post-mitotic neurons through a DNA damage response, and error-prone repair of single-strand breaks.

Huntington's disease is one of several progressive neurodegenerative disorders caused by CAG expansion in the coding region of the Huntington's disease gene (*HD*)[1–3]. Disease severity and onset depend on the number of CAG repeats, which expand in germ cells during differentiation. Recently, CAG expansion has also been detected in brains of Huntington's disease patients; large increases in CAG length are observed in striatum, the region most affected in disease[4]. Age-dependent somatic expansion in brain cells cannot be monitored in humans; however, it is well documented in mouse models for Huntington's disease[5,6] and myotonic dystrophy[7]. In R6/1 mice, which harbour a transgene containing exon 1 of the human *HD* gene and include the CAG repeat, the inherited repeat tracts are stably maintained from birth until roughly 11 weeks of age, but begin to expand at midlife and continue to increase in length as these animals age[6]. The expanding CAG tract serves as a template for synthesis of an increasingly toxic HD protein in brain[1]. Thus, in addition to the inherited expansion, somatic changes in repeat tracts may contribute to toxicity. Indeed, published experiments demonstrate that expression of the expanded *HD* gene is toxic in somatic cells, and that cell death is accelerated and directly proportional to repeat length[1,2]. These data suggest that somatic expansion may modulate onset and progression of toxicity, and that blocking somatic expansion in the brain would be beneficial. However, the mechanism by which CAG expansions might occur in post-mitotic neurons remains unclear.

Expansion correlates with DNA oxidation *in vivo*

We investigated whether age-dependent expansion in young (7–15-week-old) and older (15–52-week-old) R6/1 transgenic mice had any relationship to oxidation, a major factor in ageing (Fig. 1). Somatic tissues of R6/1 mice vary in the degree of expansion. However, tail, brain and liver were particularly informative because they represented tissues with very different degrees of age-dependent instability[6] (Fig. 1a).

We found that the level and accumulation of oxidative DNA damage correlated well with the degree of expansion (Fig. 1a, b). For example, 7,8-dihydro-8-oxoguanine (8-oxo-G) in the tail was low and expansion was modest at all ages tested, whereas in liver and in brain, the lesion level was high and expansion continued to progress with age (ref. 6; Fig. 1b). Oxidative lesions including 8-oxo-G, 5-hydroxyuracil (5-OH-uracil), 5-hydroxycytosine (5-OHC), and formamidopyrimidine (FAPY) tended to accumulate in brain and liver of R6/1 animals as they age from 7 to 52 weeks (Fig. 1c, Supplementary Fig. 1a). Neither uracil nor 3-methyladenine (3-meA) accumulated in any tissue at any age tested (Fig. 1c). Thus, the age-dependent accumulation in DNA damage seemed to be somewhat restricted to oxidative lesions. Elevation of oxidative damage was not limited to R6/1 animals. Control animals of equivalent ages accumulated the same degree of oxidative lesions in all tissues tested (Fig. 1b, c). Thus, the level of oxidation was not due to the presence of the transgene, but occurred during the process of ageing.

No reduction in repair of DNA damage in R6/1 mice

The rise in oxidative DNA damage might reflect a decrease in the capacity to repair these lesions or an increase in endogenous oxidation state with age. To distinguish between these two possibilities, we directly measured the repair activity in tissue extracts from ageing control and R6/1 animals. Repair of oxidized bases is typically initiated by cleavage of the C1 glycosidic bond by the action of a DNA glycosylase, followed by ribose-phosphate removal and generation of a single-strand break (SSB)[8]. To evaluate repair activity, we synthesized a DNA oligonucleotide containing a single base lesion, and measured generation of a 22 nucleotide cleavage product (Supplementary Fig. 1b) after incubation with tissue extracts of young or old animals (Fig. 1d, illustrated is 8-oxo-G). For 8-oxo-G, the most common lesion in DNA, we found no differences in repair activity between control or R6/1 mice at any age tested (Fig. 1d). Moreover, repair was not significantly decreased with age for other lesions tested (5-OHC, 3-meA, FAPY) with the exception of uracil (Fig. 1e). Thus, *in vivo*, accumulation of oxidative lesions did not reflect a loss of repair, but rather an increase in endogenous oxidative damage during normal ageing.

Expansion occurs during normal repair of SSB

Although trinucleotide expansion in the brain has been observed *in vivo*, it has not yet been determined whether the expansions are present in the terminally differentiated neurons in R6/1 animals. To

[1]Department of Pharmacology and Experimental Therapeutics, [2]Department of Biochemistry and Molecular Biology, [3]Neuroscience Program Mayo Clinic and Foundation, 200 First Street SW, Rochester, Minnesota 55905, USA. [4]Centre for Molecular Biology and Neuroscience and Institute of Medical Microbiology, Rikshospitalet-Radiumhospitalet HF, University of Oslo, N-0027 Oslo, Norway. [5]Laboratory of Structural Biology, National Institute of Environmental Health Sciences/National Institutes of Health, 111 TW Alexander Drive, Research Triangle Park, North Carolina 27709, USA.

address this issue, we isolated pure populations of neurons or astrocytes from the cortex of R6/1 animals using fluorescence-activated cell sorting (FACS)[9] (Supplementary Fig. 2a) and tracked CAG expansion (Fig. 2a, b). CAG somatic expansions were, indeed, present in the terminally differentiated neurons from ageing animals. The CAG tracts in pure sorted neurons were larger than those in the tail at 3 weeks, and were similar in size to the CAG repeats in whole-cortex suspension before sorting (Fig. 2a, sorted neurons versus cortex mixture, 27 weeks).

Because the accumulation of oxidative damage in the brain correlated with age-dependent expansion *in vivo*, we tested whether base oxidation could directly lead to CAG expansion. CAG repeats posed no impediment to nicking and repair *in vitro*. In the presence of protein extract and radiolabelled nucleotides, the CAG tracts within the human *HD* allele were heavily labelled when purified plasmid DNA carrying the repeats was exposed to oxidizing conditions *in vitro* (Supplementary Fig. 3).

We next tested the effects of base oxidation on the human *HD* allele in cultured cells. We treated human Huntington's disease fibroblasts with non-physiological levels of H_2O_2, to artificially induce DNA oxidation and evaluated CAG repeat length. Peroxide was chosen as the oxidizing agent because a major fraction of endogenous DNA damage arises during mitochondrial respiration when the superoxide anion radical ($O_2^{\bullet -}$) is converted into hydrogen peroxide[10]. We have previously shown in these cells that repeats are stably maintained when DNA was replicated over multiple cell divisions[11]. In the same cells, however, acute treatment with H_2O_2 uniformly led to CAG expansion of medium-length and disease-length alleles[11] (Fig. 2c–e). Indeed, peroxide treatment caused a dose-dependent increase in SSB as measured by the comet assay (Fig. 2f). Cells remained viable during the treatment (Supplementary Fig. 4) and repaired the SSBs within 2 h (Fig. 2f). Thus, oxidative DNA damage induced CAG expansion, which occurred in the process of repairing SSB.

DNA expansion in human disease is length- and sequence-dependent. To test whether oxidative damage *in vitro* caused expansion in a length- and sequence-dependent manner, we examined the stability of other repetitive sites. We failed to observe expansions at other sites. TTC/GAA repeats have been reported to be the longest tandem repeating array in the genome of humans[12]. Despite the fact that this sequence contained C and G nucleotides every third base, we were unable to detect expansion at this locus in peroxide-treated cells (Supplementary Fig. 2b). The results were similar for other sites tested (data not shown). Only CAG repeats at the long *HD* locus expanded *in vitro* under conditions of oxidative damage, whereas non-CAG sites seemed to be repaired faithfully.

Age-dependent CAG expansion *in vivo* depends on OGG1

In vivo, oxidized bases are repaired predominantly by the BER pathway[13]. The significant level of SSB in Huntington's disease fibroblasts after peroxide treatment was consistent with such a mechanism. Thus, we tested whether DNA glycosylases, the enzymes which initiate BER, contributed to CAG expansion. We first examined the role of OGG1, the primary enzyme that recognizes and removes 8-oxoG from opposite C in DNA[14,15]. We crossed R6/1 mice with mice lacking OGG1 (ref. 16) and measured the effects on age-dependent somatic expansion *in vivo* (Fig. 3a, b). Surprisingly, we found that loss of this single glycosylase significantly suppressed or delayed age-dependent expansion *in vivo* (Fig. 3a, b). Although loss of OGG1 inhibited expansion overall, the effect was not absolute (Fig. 3b). Expansion was prevented in roughly 70% of R6/1/$OGG^{-/-}$ animals. In the other 30%, tissue-specific, age-dependent expansion could be observed in either brain or liver as in age and gender matched R6/1 littermates (Fig. 3b).

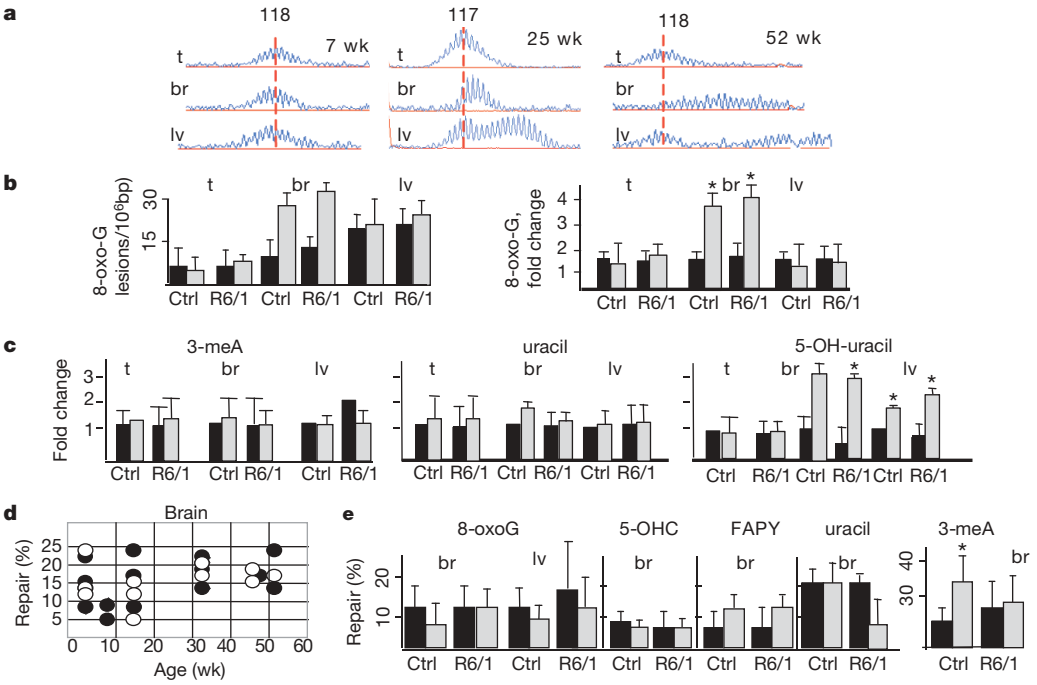


Figure 1 | Oxidative lesions accumulate in tissues of ageing mice. a, Age-dependent CAG repeat distribution in the tissues of R6/1 transgenic mice at indicated ages. The vertical line designates the midpoint length of the CAG repeat distribution in the tail of tested animals. Expansion is an increase in the number of CAG repeats indicated by the shift of distributions to the right (*x* axis is length in base pairs). **b**, Left panel, level of oxidative lesions in the tail (t), brain (br) (cortex) and liver (lv) for 8-oxo-G in control (Ctrl) and R6/1 animals at 7 (black) and 52 (grey) weeks. Right panel, accumulation (fold change) of the number of lesions from 7 to 52 weeks. Error bars, s.d. **c**, Accumulation as in **b** for 5-OH-uracil, 3-meA and uracil. *$P < 0.01$ for **b** and **c**. **d**, Repair activity (Methods) of 8-oxo-G DNA lesion in R6/1 animals (black circles) and wild-type littermate control (open circles) does not change with age (weeks) **e**, Quantified repair activity (%) of 8-oxo-G, 5-OHC, 3-meA, FAPY and uracil as in **d** for the indicated tissues at 7 (black) and 52 (grey) weeks. Reported are the mean repair activity (%) and the s.d. The limit of the s.d. is 50 (3-meA).

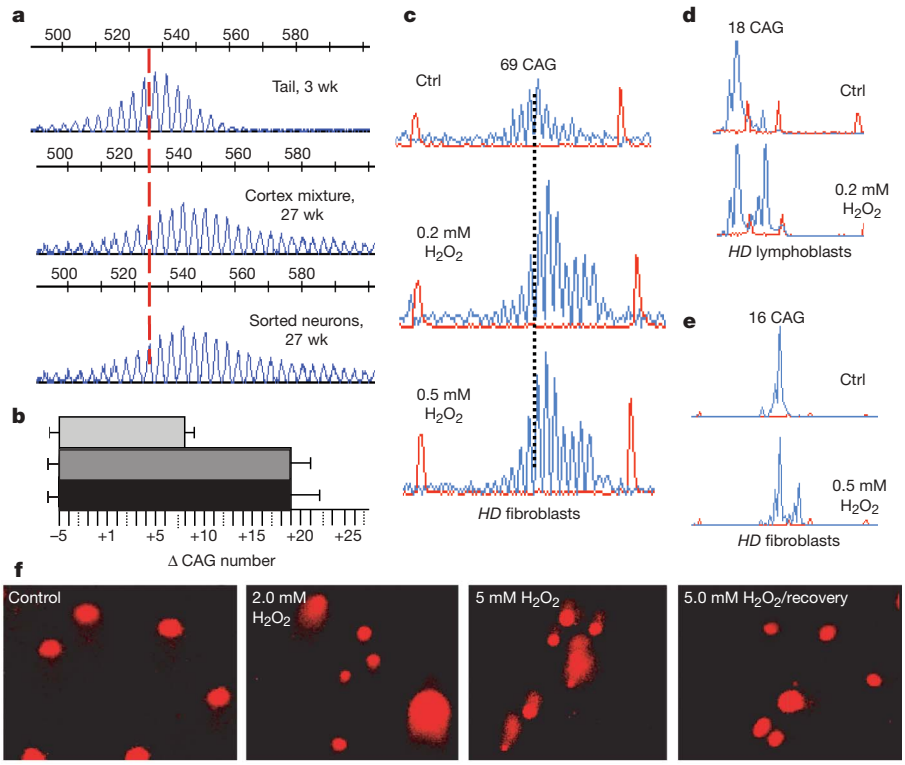


Figure 2 | Direct exposure to oxidizing agents causes expansion at the human *HD* locus *in vitro*. **a**, CAG repeat distribution of pure sorted cortical neurons at 27 weeks as compared with that of the tail at 3 weeks and to whole-cortex cell suspensions of the same animal before sorting. Vertical dashed line designates the midpoint length of the CAG repeat tract in the tail. Numbers represent size standards. **b**, The quantified sizes of the repeat tract in tail at 3 weeks (light grey), cortex mixture (dark grey) and pure sorted neurons at 27 weeks (black). Error bars, s.d.; Δ, mean length change of CAG repeats. **c–e**, Fibroblasts (**c** and **e**) and lymphoblasts (**d**) from Huntington's disease patients treated in culture with indicated concentrations of hydrogen peroxide. The expanded allele is 69 repeats and the normal allele is 16 CAG repeats. Expansion in peroxide-treated cells versus untreated cells (Ctrl) was determined as in Fig. 1. **f**, Comet assay for SSBs. Increasing peroxide treatment as indicated induces SSB as detected by comet tails. SSBs were repaired in cells by 2 h post treatment (5 mM H_2O_2/recovery) as judged by the loss of the comet tails.

Many glycosylases have preferred but overlapping substrate specificity[8,13–19]. Therefore, the dependence of CAG expansion for OGG1 was unexpected because, *in vivo*, any particular glycosylase can be 'backed-up'. To test further the specificity of OGG1 in the expansion process, we created two additional lines by crossing R6/1 mice with mice lacking alkyladenine glycosylase (AAG)[20] or NTH1 (homologue of *Escherichia coli* endonuclease III)[21] (Supplementary Fig. 5a). NTH1 prefers to excise thymine glycol[17,19,21] whereas AAG excises a variety of

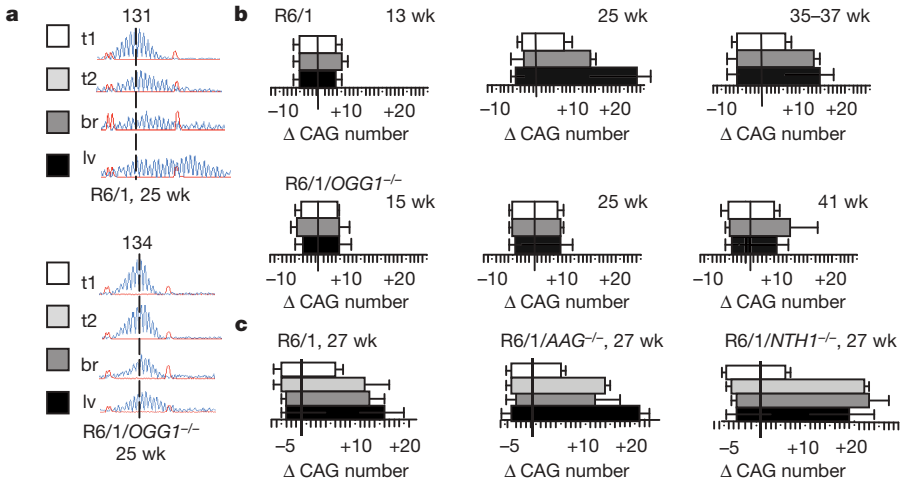


Figure 3 | Age-dependent expansion is suppressed in mice lacking OGG1 glycosylase. **a**, **b**, Representative CAG repeat distributions in the tissues of R6/1 transgenic or R6/1/*OGG1*$^{-/-}$ mice. **a**, Vertical dashed line designates the midpoint of the CAG repeat tract length in the tail at 3 weeks (t1, white box). The CAG tract lengths in the tail (t2, light grey), brain (br, dark grey) (cortex) and liver (lv, black) at 25 weeks in the same animal are shown. **b**, The mean length change of CAG repeats (Δ) at the indicated ages in the tissues of R6/1 mice and R6/1/*OGG1*$^{-/-}$ mice. Each value is expressed as the mean change and the s.d. Tissues are indicated by colour: tail at 3 weeks in white; brain and liver at indicated ages in dark grey and in black respectively. **c**, Quantified data for mean length change of CAG repeats from R6/1 mice, R6/1/*AAG*$^{-/-}$ or R6/1/*NTH*$^{-/-}$. Tissues are indicated by colour: tail at 3 weeks in white; tail, brain and liver at 27 weeks in light grey, dark grey and black respectively. Analysis is the same as in **a**, **b**.

alkylated bases with the highest affinity for 3-methyladenine[18,20]. In contrast to R6/1/*OGG1*$^{-/-}$ animals, deletion of either AAG or NTH1 in R6/1/*AGG*$^{-/-}$ and R6/1/*NTH1*$^{-/-}$ mice did not reduce somatic expansion relative to their R6/1 littermate controls at any age tested (Fig. 3c; Supplementary Fig. 5a). Thus, loss of the single glycosylase, OGG1, was a dominant factor contributing to age-dependent expansion *in vivo*, despite the fact that other DNA glycosylases were present and the tissues were competent for SSB repair.

Somatic expansion in R6/1 mice is length- and sequence-dependent

We tested whether somatic expansion *in vivo* in R6/1 animals recapitulated the features that characterize the mutation in humans. First, it is widely accepted that the expansion mutation occurs primarily at repeats capable of forming secondary structure[2,3,22–24], and is not accompanied by general microsatellite instability[2,25] Second, in human disease, there is a threshold CAG length, below which there is little probability of observing expansion. In humans, CAG expansion is typically observed at tract lengths above 36 repeats[26,27], whereas in mouse, the threshold is around 100 (refs 5, 6, 28). Therefore, we tested *in vivo* whether age-dependent expansion in R6/1 mice was (i) length-dependent, (ii) sequence-dependent, (iii) occurred at other repetitive sequences, and (iv) whether loss of OGG1 altered these properties.

Expansion did not occur at short CAG repeats in R6/1 and in R6/1/*OGG1*$^{-/-}$ animals—it was not observed in endogenous mouse *HD* gene (6 CAG repeats) in either line (not shown). We next tested unrelated alleles containing longer CAG tracts. We found that CAG tracts on chromosome 9 (19 repeats) and chromosome 7 (31 repeats) were stable in both R6/1 animals and in R6/1/*OGG1*$^{-/-}$ mice (Supplementary Fig. 5b). Thus, loss of OGG1 *in vivo* blocked expansion only at the long CAG tract (135 repeats) within the human *HD* transgene, the length of which was above the threshold reported for the mouse homologue. We observed no age-dependent expansion in R6/1 and in *R6/1/OGG1*$^{-/-}$ animals at sequences that lacked structure-forming capability. Shown is polyA microsatellite (Supplementary Fig. 5b) but the same result was observed for other microsatellites tested (data not shown). Thus, age-dependent somatic expansion in R6/1 mice showed the properties observed in human disease.

OGG1 initiates CAG expansion during BER reconstituted *in vitro*

The *in vivo* requirement of OGG1 for expansion of CAG repeats implicated a BER mechanism. Therefore, we tested whether expansion could be regenerated *in vitro* using purified human OGG1 and the BER machinery. DNA polymerases can generate expansion on CAG-containing templates by primer extension[29–32] and on substrates that mimic BER intermediates[32,33]. However, no experiments have tested the entire BER process beginning with the glycosylase incision of a lesion within CAG duplex DNA. We synthesized two DNA templates, each 100 base pairs in length, in which a single 8-oxoG base was positioned 23 nucleotides from the 5′ end of one strand (Supplementary Fig. 6a, b, top panels). In the random template control (Supplementary Fig. 6a), the base lesion was flanked by random sequence DNA of roughly equal CG/AT content, whereas in the CAG template, the sequence was identical except that the base lesion was flanked on the 3′ side by 19 CAG repeats (Supplementary Fig. 6b).

We evaluated the BER pathway using a step-by-step addition of OGG1, apurinic/apirimidinic endonuclease (APE1) and polymerase β (Polβ) to the reconstituted *in vitro* system. Removal of the damaged base by OGG1/APE cleavage generates a fragment of 22 nucleotides (Supplementary Fig. 6a, b). Polβ was chosen as the gap-filling polymerase because it is the major BER polymerase in mammalian cells, and because it favours single nucleotide additions[34]. If the CAG sequences did not promote expansion, the CAG templates should behave as a random template. Polβ would restore the initial 23 nucleotide length and the template to 100 nucleotides after ligation (Supplementary Fig. 6a). On the other hand, if the CAG sequences promoted expansion, then some strand displacement/slippage should occur. In this case, Polβ should yield multi-nucleotide additions ($n > 1$ nucleotide) and templates longer than 100 nucleotides following ligation of looped intermediates (Supplementary Fig. 6b).

With both templates, OGG1 activity was strong, and no difference was observed in the products between the CAG and random templates (Fig. 4a, lanes 1 and 4). Addition of APE1 resulted in nicking of the phosphodiester backbone and the production of the 22 nucleotide fragment (Fig. 4a, lanes 2 and 5). In the absence of APE, Polβ was unable to carry out strand extension of the 22 nucleotide fragment, as expected (data not shown).

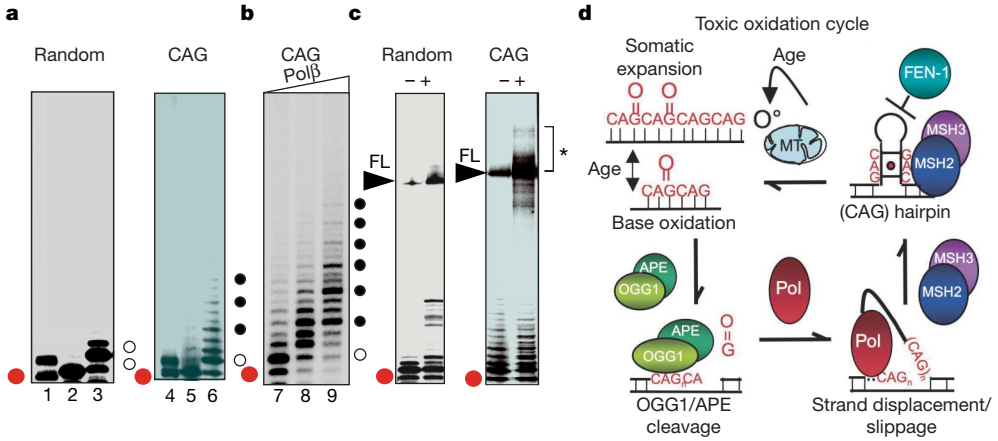


Figure 4 | OGG1 excision of 8-oxo-G within CAG repeat DNA can initiate strand displacement and expansion *in vitro* during BER. a, Reaction products after step-wise addition of OGG1 (lanes 1 and 4), OGG1 + APE (lanes 2 and 5), and OGG1 + APE + Polβ (lanes 3 and 6) to random (left) or CAG (right) templates. Red dots indicate 22 nucleotide excision/incision product, open and filled circles are 1 nucleotide and 3 nucleotide additions, respectively. **b**, Triplet pattern on CAG template as a function of increasing Polβ concentration indicated by the triangle: 0 nM (lane 7); 0.5 nM (lane 8); and 1 nM (lane 9). **c**, Products of the BER reactions for random (left) or CAG (right) templates in the absence (−) or presence (+) of DNA ligase. FL, 100 nucleotide full-length product after ligation. The bracket and asterisk depict the 100 nucleotide FL template and larger expansion products. Red dot as in **a**. **d**, 'Toxic oxidation cycle' model for age-dependent somatic expansion. Endogenous oxidative radicals (O°) arising from mitochondrial (MT) respiration creates oxidative DNA lesions. Under conditions of normal BER, OGG1/APE cleavage produces a nick, and polymerase (Pol) facilitates hairpin formation during gap-filling synthesis. CAG hairpins are stabilized by MSH2/MSH3 binding (red dot is a mismatch in the stem) and escape FEN-1 loading and cleavage owing to a hidden 5′ end. The hairpin intermediate is processed to restore duplex DNA generating a longer CAG template, which is again subject to oxidative DNA damage. The cycle continues with age.

We found that gap filling by Polβ on the random template resulted in a single nucleotide addition as the major product, and restored the base at the position 23 nucleotides from the end (Fig. 4a, lane 3). In contrast, on CAG templates, OGG1/APE cleavage and gap filling synthesis by Polβ generated longer addition products (Fig. 4a, lane 6; black dots). Three nucleotide additions were favoured (Fig. 4b, black dots). On random templates, the addition of ligase primarily restored the 100 nucleotide full-length starting material (Fig. 4c, lane 2), whereas on the CAG template, ligation resulted in appearance of fragments corresponding to the starting material as well as expanded products (shown in starred bracket Fig. 4c). Thus, OGG1-mediated BER was able to initiate expansion through strand displacement/slippage during the gap-filling step of BER, and both displacement and expansion depended on the CAG sequence.

Discussion

Here we demonstrate that, *in vivo*, age-dependent somatic CAG expansion is initiated in the process of removing oxidized DNA bases, and is dependent on a single DNA glycosylase, OGG1. The unexpected specificity for OGG1 is not due to a specific rise of its preferred 8-oxo-G lesions because many oxidative DNA lesions accumulate in the brains of ageing R6/1 animals. Further, the requirement for OGG1 in CAG expansion does not seem to arise from lack of functional back-up systems. All DNA glycosylases are present in R6/1 animals and the tissues were competent for lesion repair (Fig. 1, Supplementary Fig. 1). Because only OGG1 suppressed the mutation, we favour a model in which the unusual dependence of expansion on OGG1 arises from its interaction with the CAG tract in a manner that is not shared by other glycosylases. Thus, redundant biochemical activities for removing oxidative DNA damage *in vitro* may promote distinct biological consequences *in vivo*, which are relevant to the mechanism for human neurodegenerative disease.

The specificity of OGG1 may indicate that 8-oxoG lesions are favoured within CAG sites and occur frequently. Alternatively, OGG1 might bind better to CAG sequences or their DNA tertiary structures, or binding may induce conformations that confer specificity. CAG sequences can form DNA hairpins with mismatched bases every third nucleotide. We have previously shown that binding of the MSH2/MSH3 complex to the CAG hairpin alters properties needed for recognition and repair of mispaired bases[35]. Moreover, we and others have demonstrated that loss of mismatch repair proteins, MSH2 (refs 6, 35–37), MSH3 (refs 35, 38) and PMS2 (ref. 39) abrogates age-dependent somatic expansion *in vivo* in a number of mouse models for Huntington's disease and myotonic dystrophy. Because loss of OGG1 also suppresses the mutation, these data imply that the mismatch repair complex physically or functionally cooperates with OGG1 to promote expansion of the CAG repeat sequences.

Although many models have been proposed for trinucleotide expansion on the basis of *in vitro* systems, the results presented here support the hypothesis that, *in vivo*, somatic expansion in mammalian cells initiates from an OGG1-mediated BER mechanism. Somatic expansion does not require cell division. We find that somatic age-dependent expansion occurs in neurons well after these cells are terminally differentiated and mitotic replication has ceased. Likewise, it has been demonstrated that loss of RAD52, RAD54 or Ku did not suppress CTG expansion in the mouse model for myotonic dystrophy[36]. Thus, enzymes generally needed for repair of double-strand breaks are not required for expansion *in vivo*. Our data indicate that a SSB mechanism such as BER is more likely. Highly processive polymerases, such as Polδ or Polε, can copy thousands of base pairs, and are often used to complete long-patch BER[40]. Although we have not yet tested this hypothesis, strand displacement and gap-filling synthesis over a long stretch of triplet repeat DNA might facilitate formation of larger structural intermediates. Thus, long patch BER might also be relevant for large expansions associated with non-coding regions of genes[1,2].

Inheritance of an expanded CAG tract is the underlying cause for Huntington's disease toxicity. However, we propose a 'toxic oxidation cycle' model in which somatic mutations contribute to onset and progression. The oxidation–excision–expansion cycle escalates with age as oxidative lesions in the brain accumulate (Fig. 4d). The increase in oxidative lesions elicits a DNA damage response and increases the need for OGG1 repair through a SSB mechanism. Expansion would result from 'error-prone repair' in the process of gap-filling synthesis by a strand displacement/slippage mechanism (Fig. 4d). Loop entrapment on one strand can place constraints on the other strand and initiate restoration of duplex DNA. Because the 'toxic oxidation cycle' is iterative with age, the resulting mutation would be predicted to grow progressively (Fig. 4d). Oxidative damage has been recognized as a major cause of ageing and neurological disease in all aerobic organisms[11,14]. CAG expansion at the disease allele in Huntington's disease provides a glimpse of the underlying molecular mechanism for at least one neurodegenerative disease. However, the accumulation of oxidized bases in the brain overwhelms the repair machinery throughout the genome, thereby increasing the probability of unrepaired SSBs, which can be fatal[41]. Thus, the importance of oxidative DNA damage and SSB intermediates probably extends beyond Huntington's disease and has general relevance to late onset neurodegeneration.

METHODS SUMMARY

Transgenic male mice B6CBA-TgN R6/1 (ref. 5) were crossed with C57BL/6J female partners that lacked one of the glycosylases—AAG (ref. 20), OGG1 (ref. 16) or NTH1 (gift from L. Samson). Cell lines from Huntington's disease patients used in the study are described in Methods. For the comet assay, SK-N cells untreated or treated with H_2O_2 were used. For *in vitro* oxidation, a plasmid DNA that contained a truncated version of human *HD* complementary DNA with 40 CAG repeats[42] was treated with methylene blue, exposed to visible light and used as a substrate in *in vitro* repair assay. Statistical analysis for CAG repeat change was performed, and change was expressed as mean ± standard deviation (s.d.). Statistical significance was determined by two-way analysis of variance (ANOVA), followed by Fisher's exact test. Details of FACS analysis, preparation of nuclear extracts and enzyme cleavage assays, *in vitro* reconstitution of BER, preparation of nuclear DNA and lesion detection by HPLC can be found in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 August 2006; accepted 2 April 2007.
Published online 22 April 2007.

1. Cummings, C. J. & Zoghbi, H. Y. Trinucleotide repeats: mechanisms and pathophysiology. *Annu. Rev. Genom. Hum. Genet.* **1,** 281–328 (2000).
2. Pearson, C. E., Edamura, K. N. & Cleary, J. D. Repeat instability: mechanisms of dynamic mutations. *Nature Rev. Genet.* **10,** 729–742 (2005).
3. Kovtun, I. V. & McMurray, C. T. *Genetic Instabilities and Hereditary Neurological Diseases* (ed. Wells, R.D.) 679–690 (Academic Press, 2006).
4. Kennedy, L. *et al.* Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* **12,** 3359–3367 (2003).
5. Mangiarini, L. *et al.* Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nature Genet.* **15,** 197–200 (1997).
6. Kovtun, I. V. & McMurray, C. T. Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genet.* **27,** 407–411 (2001).
7. Seznec, H. *et al.* Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum. Mol. Genet.* **9,** 1185–1194 (2000).
8. Dizdaroglu, M. Substrate specificities and excision kinetics of DNA glycosylases involved in base-excision repair of oxidative DNA damage. *Mutat. Res.* **531,** 109–126 (2003).
9. Sergent-Tanguy, S., Chagneau, C., Neveu, I. & Naveilhan, P. Fluorescent activated cell sorting (FACS): a rapid and reliable method to estimate the number of neurons in a mixed population. *J. Neurosci. Methods* **129,** 73–79 (2003).
10. Brand, M. D. *et al.* Mitochondrial superoxide: production, biological effects, and activation of uncoupling proteins. *Free Radic. Biol. Med.* **37,** 755–767 (2004).
11. Kovtun, I. V., Thornhill, A. R. & McMurray, C. T. Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. *Hum. Mol. Genet.* **13,** 3057–3068 (2004).

12. Chauhan, C., Dash, D., Grover, D., Rajamani, J. & Mukerji, M. Origin and instability of GAA repeats: insights from Alu elements. *J. Biomol. Struct. Dyn.* **20,** 253–263 (2002).
13. Bjelland, S. & Seeberg, E. Mutagenicity, toxicity and repair of DNA base damage induced by oxidation. *Mutat. Res.* **531,** 37–80 (2003).
14. Auffret van der Kemp. P., Thomas, D., Barbey, R., de Oliveira, R. & Boiteux, S. Cloning and expression in *Escherichia coli* of the *OGG1* gene of *Saccharomyces cerevisiae,* which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamido pyrimidine. *Proc. Natl Acad. Sci. USA* **93,** 5197–5202 (1996).
15. Zharkov, D. O., Rosenquist, T. A., Gerchman, S. E. & Grollman, A. P. Substrate specificity and reaction mechanism of murine 8-oxoguanine-DNA glycosylase. *J. Biol. Chem.* **275,** 28607–28617 (2000).
16. Klungland, A. *et al.* Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proc. Natl Acad. Sci. USA* **96,** 13300–13305 (1999).
17. Dizdaroglu, M., Karahalil, B., Senturker, S., Buckley, T. J. & Roldan-Arjona, T. Excision of products of oxidative DNA base damage by human NTH1 protein. *Biochemistry* **38,** 243–246 (1999).
18. O'Brien, P. J. & Ellenberger, T. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.* **279,** 9750–9757 (2003).
19. Marenstein, D. R. *et al.* Substrate specificity of human endonuclease III (hNTH1). Effect of human APE1 on hNTH1 activity. *J. Biol. Chem.* **278,** 9005–9012 (2003).
20. Engelward, B. P. *et al.* Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase. *Proc. Natl Acad. Sci. USA* **94,** 13087–13092 (1997).
21. Hazra, T. K. *et al.* Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl Acad. Sci. USA* **99,** 3523–3528 (2002).
22. Kovtun, I. V., Goellner, G. & McMurray, C. T. Structural features of trinucleotide repeats associated with DNA expansion. *Biochem. Cell Biol.* **279,** 325–336 (2001).
23. Gacy, A. M. *et al.* GAA instability in Friedreich's Ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol. Cell* **1,** 583–593 (1998).
24. Spiro, C. *et al.* Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell* **4,** 1079–1085 (1999).
25. Goellner, G. M. *et al.* Different mechanisms underlie DNA instability in Huntington disease and colorectal cancer. *Am. J. Hum. Genet.* **60,** 879–890 (1997).
26. Kremer, B. *et al.* Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. *Am. J. Hum. Genet.* **57,** 343–350 (1995).
27. Chong, S. S. *et al.* Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analysis. *Hum. Mol. Genet.* **6,** 301–309 (1997).
28. Wheeler, V. C. *et al.* Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum. Mol. Genet.* **8,** 115–122 (1999).
29. Wilson, S. H. *et al.* in *Genetic Instabilities and Hereditary Neurological Diseases* (eds Wells R.D. & Warren, S.T.) 493–698 (Academic Press, 1998).
30. Petruska, J., Hartenstine, M. J. & Goodman, M. F. Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *J. Biol. Chem.* **273,** 5204–5210 (1998).
31. Hartenstine, M. J., Goodman, M. F. & Petruska, J. Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase. *J. Biol. Chem.* **275,** 18382–18390 (2000).
32. Hartenstine, M. J., Goodman, M. F. & Petruska, J. Weak strand displacement activity enables human DNA polymerase β to expand CAG/CTG triplet repeats at strand breaks. *J. Biol. Chem.* **277,** 41379–41389 (2002).
33. Lyons-Darden, T. & Topal, M. D. Abasic sites induce triplet-repeat expansion during DNA replication *in vitro*. *J. Biol. Chem.* **274,** 25975–25978 (1999).
34. Beard, W. A. & Wilson, S. H. Structure and mechanism of DNA polymerase β. *Chem. Rev.* **106,** 361–382 (2006).
35. Owen, B. A. *et al.* $(CAG)_n$-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nature Struct. Mol. Biol.* **12,** 663–670 (2005).
36. Savouret, C. *et al.* CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.* **22,** 2264–2273 (2003).
37. Manley, K., Shirley, T. L., Flaherty, L. & Messer, A. MSH2 deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genet.* **23,** 471–473 (1999).
38. van den Broek, W. J. *et al.* Somatic expansion behaviour of the $(CTG)_n$ repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet.* **11,** 191–198 (2002).
39. Gomes-Pereira, M., Fortune, M. T., Ingram, L., McAbney, J. P. & Monckton, D. G. Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum. Mol. Genet.* **13,** 1815–1825 (2004).
40. Stucki, M. *et al.* Mammalian base excision repair by DNA polymerases δ and ε. *Oncogene* **17,** 835–843 (1998).
41. Ahel, I. *et al.* The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature* **443,** 713–716 (2006).
42. Trushina, E. *et al.* Microtubule destabilization and nuclear entry are sequential steps leading to toxicity in Huntington's disease. *Proc. Natl Acad. Sci. USA* **100,** 12171–12176 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Mayo Foundation, the National Institutes of Health (C.T.M.), and, in part, by the Intramural Research Program of the National Institute of Environmental Health Sciences. The authors wish to dedicate this work to Erling Seeberg. We thank J. Hoeijmakers for $NTH^{-/-}$ and L. Samson for $AAG^{-/-}$ mice, and N. Kinzel for help in cortex dissection.

Author Contributions C.T.M. oversaw the entire project. I.V.K. and C.T.M. conceived the experiments, wrote the manuscript, and prepared all Figures. I.V.K. carried out *in vitro* and *in vivo* experiments in cell lines and in mice (animal breeding, oxidation, comet assay, FACS, analysis of repeat size, dissected animal tissue) and prepared tissue extracts for all testing of repair activity *in vitro*, and performed analysis of the results. Y.L. with S.H.W. supervision performed base excision repair reconstitution experiments, M.B. and A.K. carried out DNA repair assays. All co-authors contributed to the manuscript with their comments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.T.M. (mcmurray.cynthia@mayo.edu).

METHODS

Mouse lines and breeding. Transgenic male mice B6CBA-TgN R6/1 (ref. 5) were crossed with C57BL/6J female partners that lacked one of the glycosylases—AAG (ref. 20), OGG1 (ref. 16) or NTH1 (gift from L. Samson)—and bred until they were homozygous knockout. Litters were screened for the presence of human *HD* transgene[6] (see below) and the absence of each glycosylase by PCR as described[16,20].

Cell cultures and treatments. Human Huntington's disease fibroblasts (GM04281, GM04687, GM04212; Coriell Cell Repositories) and human Huntington's disease lymphoblasts (18/49 CAG, 17/69 CAG), kindly provided by E. Almqvist, were maintained in MEM medium with 20% of FBS. For the treatment, cells were grown until they reached 70% confluence and were treated with indicated concentrations of H_2O_2 for 30 min. Cells then were washed, medium was replaced, cells were allowed to recover for 3 days and treated again 1 or 2 times in the same manner. After third treatment cells were collected and DNA was isolated. Lymphoblasts were collected by centrifugation between treatment and washing. CAG repeat sizing were performed as described below.

Comet assay. SK-N cells untreated or treated with H_2O_2 were scraped off the plates, pelleted, re-suspended in PBS (10^4 cells per 10 µl), mixed with low melting agarose (at 37 °C) and mounted on agarose-coated (0.5% in PBS) slides. Cells were lysed in a buffer containing 10 mM Tris (pH 10), 2.5 M NaCl, 100 mM EDTA, 1% triton X-100 at 4 °C for 1 h and subjected to electrophoresis in 300 mM NaOH, 1 mM EDTA buffer. Slides were neutralized in 0.4 M Tris (pH 7.5) and stained with ethidium bromide (20 mg ml^{-1}). Individual cells were imaged using a LSM 510 (Zeiss) microscope with ×20 objective. Fluorescence was visualized by using Ar/Kr laser at 488/510 nm (excitation/emission).

***In vitro* oxidation of CAG repeat-containing plasmid.** Plasmid DNA that contained truncated version of human *HD* cDNA with 40 CAG repeats (cDNA was inserted into *Not*1 site of pEGFP-C1 vector; Clontech)[42] was treated with methylene blue (10 µM in phosphate buffer, pH 7.4) and exposed to visible light for 8 min. Ethanol precipitated and recovered plasmid then, was used as a template for *in vitro* repair reaction. Protein extract from PC-12 cells was incubated with the oxidized plasmid DNA in the repair buffer (50 mM HEPES, 5 mM $MgCl_2$, 2 mM DTT, 0.2 mM EDTA) with the addition of 40 mM phosphocreatine, creatine kinase, dATP, dTTP, dGTP[32] and dCTP[32] for 1.5 h. Repair products were precipitated, digested with *Not*1 to isolate CAG repeat tract and resolved on denaturing PAGE gel.

CAG repeat sizing. The DNA was purified from mouse tissues or cultured human fibroblasts[6] and used for PCR amplification of various microsatellites. CAG repeats at the human *HD* locus were amplified as described previously[5,6]. For amplification of GAA repeats on human chromosome 5 (accession no. AC011416) 5′-CCTTCTGTCTTAC TTCATAG-3′ and 5′-CAGCAAAGTG-TGTGTGTGGTT-3′ primers were used. CAG repeats on chromosome 9 (accession no. AC100550) and chromosome 7 (accession no. AC122399) in the mouse genome were amplified using primers: set 1, 5′-CTCTGCACTGTGTTCAGG-GAC-3′ and 5′-ACTGATGCAGCCCAGGTACTG-3′; and set 2, 5′-GGAA-GGACCTTCATAGGCTTCT-3′ and 5′-TGCCTATCTTATCCAGCTAGGC-3′ respectively. The polyA locus at chromosome 17 (accession no. AC096777) was amplified as described[43]. One primer in each PCR amplification was labelled with FAM-6, and fluorescent PCR products were analysed using ABI prism 3700 DNA analyser instrument and the GeneMapper software v3.

Preparation of nuclear extracts and enzyme cleavage assays. Organs were removed, dissected, quick-frozen in liquid nitrogen, and stored at 80 °C until use. Nuclear extracts were prepared as described previously[44].

DNA repair assays. The assay for 8-oxoG-, uracil- and 5-OHC-containing DNA fragments was performed as follows. Duplex DNA substrates containing the indicated lesions at position 23 were generated by ^{32}P-end labelling the 5′-end of the 49-mer oligonucleotides (5'-GGCGGCATGACCC [8-oxoG or uracil] GAGGC CCATC-3′) by T4 polynucleotide kinase (MBI Fermentas) and [γ-^{32}P] ATP (3,000 Ci mmol^{-1}; Amersham). The labelled oligonucleotides were annealed to complementary strands with cytosine and guanine opposite to 8-oxoG or uracil, respectively. Similarly, a 49-bp-long duplex DNA substrate containing a single 5-OHC at indicated position (5-AATTGCGATCT-AGCTCGCCAG- [5-OHC] AGCGACCTTATCTGATGA-3′) was 5′-^{32}P-end-labelled as described above and hybridized to a complementary strand with guanine opposite to 5-OHC. The enzyme activities were assayed in a reaction buffer containing 25 mM 3-(*N*-morpholino) propanesulphonic acid (MOPS), pH 7.5, 0.5 mM DTT, 0.5 mM EDTA and 2.5% glycerol for 30 min at 37 °C. Reaction mixtures contained 100–500 fmol substrate and 2 ug of protein extract in a total volume of 10 µl. The products of the reactions were analysed by 20% denaturing polyacrylamide gel electrophoresis and phosphorimaging.

FAPY and 3-MeA DNA glycosylase activity assays were performed as follows. *N*-[H^3]methyl-*N*′-nitrosourea (MNU; 1.5 Ci mmol^{-1}) was used to prepare alkylated calf thymus DNA (6,000 dpm µg^{-1} DNA). 3-MeA DNA glycosylase activity was assayed in reaction buffer (70 mM MOPS, pH 7.5, 1 mM DTT, 1 mM EDTA, 5% glycerol) with 7 µg DNA substrate and 2 µg protein extract for 30 min at 37 °C in a total reaction volume of 50 µl. *N*-[^{3}H]methyl-*N*′-nitrosourea (18 Ci mmol^{-1}) was used to prepare poly(dG–dC) DNA containing FAPY residues (5,000 dpm µg^{-1} DNA) as described[44]. FAPY DNA glycosylase activity was assayed in a reaction buffer (50 mM MOPS, pH 7.5, 1 mM DTT, 1 mM EDTA and 5% glycerol) with 0.4 µg FAPY substrate and 2 µg protein extract for 30 min at 37 °C in a total volume of 50 µl. Following DNA precipitation of the reaction mixture, base removal was quantified as radioactivity in the supernatant by using a Liquid Scintillation Counter (Tri-Carb 2900TR, Packard).

Tissue preparation and fluorescent-activated cell sorting (FACS) analysis. Adult mice were euthanized and the cortex was dissected and immediately immersed into HBS (10 mM HEPES, pH 7.3, 145 mM NaCl, 22 mM KCl and 5 mM glucose). The tissue was sliced and placed in papain solution (2 mg ml^{-1} in HBS) for 30 min to dissociate cells. Cell suspension then was triturated 20–30 times with a 1 ml pipette, filtered through 70 µm mesh (Falcon), centrifuged at 900 r.p.m. at 4 °C and resuspended in PBS with BSA (1 mg mg^{-1}). Cells were fixed in 2% paraformaldehyde for 30 min on ice, washed with PBS and permeabilized by incubation in 0.1% Tween for 30 min on ice. To distinguish glia from neurons, we stained cells with cell-specific antigens. Dispersed cells were incubated with primary antibodies either against neurons-specific β III tubulin (Promega) or glial-specific GFAP (Sigma) (1/500 dilution) for 45 min on ice. Cells then were washed and incubated for 45 min on ice with FITC-conjugated anti-mouse IgG (1/1000 dilution). The labelled cells were isolated using FACS, which has been effectively used to characterize and separate different cell types derived from central nervous system. Cell discrimination was gated by both scatter and fluorescence, and only cells with a scattering profile indicative of a single cell were sorted for FITC signal. Sorting was performed using a FACSVantage (Becton Dickinson) flow cytometer. As a negative control, cells were taken through the same labelling procedure omitting the primary antibody. In that case, no cells yielded a specific FITC signal. β-III tubulin is an abundant cytoskeletal protein, and variablity in the FITC intensity results from variations of expression level, cell permeability and the number of FITC-conjugated anti-mouse IgG bound in each cell.

Preparation of nuclear DNA and analysis of 8-oxoG by HPLC–electrochemical detection (HPLC–ECD). Male mice were euthanized at indicated ages; livers were removed, quick-frozen in liquid nitrogen, and stored at 80 °C until use. Scissor-macerated liver was passed through a 19-gauge, 1.5-inch needle. Extraction of DNA and hydrolysis to nucleosides was done by nuclease P1 and alkaline phosphatase treatment. To reduce oxidation during the preparation of DNA, TEMPO (2,2,6,6-tetramethylpiperidine-*N*-oxyl) was added to all solutions at 100 µM immediately before use. 8-hydroxy-2'-deoxyguanosine and 2'-deoxyguanosine were separated by HPLC and analysed by electrochemical detection (ECD; +300 mV) and ultraviolet light (290 nm), and results were expressed as the ratio of 8-oxoG per 106 bp in each DNA sample. The following conversions were used: 1 8-oxoG per 105 G = 4 8-oxoG per 106 bp; 1 8-oxoG per 106 bp = 6,000 8-oxoG per diploid genome.

***In vitro* assays for base excision repair.** DNA oligonucleotides containing 8-oxoG were synthesized and supplied after purification by PAGE (Operon Biotechnologies). All other oligonucleotides were synthesized and purified by Integrated DNA Technologies. The 8-oxoG-containing substrates were constructed by annealing the damaged strand to its template strand at a molar ratio of 1:1.5. The substrates were radiolabelled at the 5′-end of the damaged strand using [γ-^{32}P]ATP and Optikinase (USB Corporation). Unincorporated radiolabelled material was removed with a G-25 spin column (GE Healthcare). Purified human GST-tagged OGG1, APE and Polβ, at various concentrations, were employed in BER/gap-filling reactions. Purified T4 DNA ligase (60 U µl^{-1}) (New England Biolabs) was used in some reaction mixtures as specified. The reaction mixtures contained 50 mM Tris-HCl, pH 7.5, 50 mM KCl, 0.1 mg ml^{-1} BSA, and 0.01% NP-40. Five mM $MgCl_2$ was included in APE, Polβ and ligase reaction mixtures. Fifty µM dNTPs were used for Polβ DNA synthesis and 1 mM ATP was included in the ligation reaction mixtures. Ten nM oligonucleotide substrates were used in the final reaction mixture, and these substrates had been pre-incubated with 30 nM OGG1 at 37 °C for 30 min. The final reaction mixtures (20 µl) were assembled by mixing the enzymes with the OGG1-cleaved substrates at 0 °C, and then the reaction mixtures were incubated at 37 °C for 20 min. The reactions were terminated by addition of EDTA. The substrates and products were subjected to separation through a 15–18% polyacrylamide denaturing gel containing 7 M urea. The products were detected by PhorsphorImager and quantified by ImageQuant.

Statistical analysis. The changes in CAG repeat number were quantified using simple descriptive statistics (mean change ± s.d.). Statistical significance was determined by two-way analysis of variance (ANOVA), followed by Fisher's exact test (Fig. 1).

43. Kabbarah. O. *et al.* A panel of repeat markers for detection of microsatellite instability in murine tumors. *Mol. Carcinog.* **38,** 155–159 (2003).
44. Boiteux, S., Belleney, J., Roques, B. P. & Laval, J. Two rotyameric forms of open ring 7-methylguanine are present in alkylated polynucleotides. *Nucleic Acids Res.* **12,** 542–5439 (1984).

ARTICLES

Atomic structures of amyloid cross-β spines reveal varied steric zippers

Michael R. Sawaya[1], Shilpa Sambashivan[1], Rebecca Nelson[1], Magdalena I. Ivanova[1], Stuart A. Sievers[1], Marcin I. Apostol[1], Michael J. Thompson[1], Melinda Balbirnie[1], Jed J. W. Wiltzius[1], Heather T. McFarlane[1], Anders Ø. Madsen[2,3], Christian Riekel[3] & David Eisenberg[1]

Amyloid fibrils formed from different proteins, each associated with a particular disease, contain a common cross-β spine. The atomic architecture of a spine, from the fibril-forming segment GNNQQNY of the yeast prion protein Sup35, was recently revealed by X-ray microcrystallography. It is a pair of β-sheets, with the facing side chains of the two sheets interdigitated in a dry 'steric zipper'. Here we report some 30 other segments from fibril-forming proteins that form amyloid-like fibrils, microcrystals, or usually both. These include segments from the Alzheimer's amyloid-β and tau proteins, the PrP prion protein, insulin, islet amyloid polypeptide (IAPP), lysozyme, myoglobin, α-synuclein and β_2-microglobulin, suggesting that common structural features are shared by amyloid diseases at the molecular level. Structures of 13 of these microcrystals all reveal steric zippers, but with variations that expand the range of atomic architectures for amyloid-like fibrils and offer an atomic-level hypothesis for the basis of prion strains.

Amyloid diseases are accompanied by the deposition of elongated, unbranched protein fibrils. For pathologists to designate a disease as amyloid, the fibrils must be deposited extracellularly, and must bind the dye Congo red, giving an 'apple-green' birefringence[1]. As of 2005, Alzheimer's disease and some 24 others have been found to satisfy this stringent definition[1,2]. In addition, many other proteins have been found to form amyloid-like fibrils with biophysical properties in common with amyloid fibrils. These properties include an elongated morphology[3], binding of Congo red, formation from their constituent protein molecules with cooperative, nucleation-dependent kinetics[4], and the so-called cross-β X-ray diffraction pattern[5–7]. This pattern consists of an X-ray reflection at ~4.8 Å resolution along the fibril direction, and another X-ray reflection at ~8–11 Å resolution perpendicular to the fibril direction[5,8,9]. The pattern reveals that the fibrils contain β-sheets parallel to the fibril axis, with their extended protein strands perpendicular to the axis.

Finding atomic-level structures for cross-β spines has been impeded by the fibrillar nature of amyloid, but the corner has been turned. Recent progress has included models for fibrillar segments constrained by chemical labelling, scanning proline mutagenesis, electron paramagnetic resonance, NMR, H/D exchange and X-ray fibre diffraction data[10–21], and two atomic-resolution microcrystal structures of the fibril-forming segments GNNQQNY and NNQQNY[22] from the yeast prion protein Sup35. Here we extend atomic-resolution crystallographic studies to other fibril-forming segments taken from disease-related proteins, and we relate the segments to the fibrils formed by their parent proteins.

Peptide microcrystals and fibrils

Our study of the overlapping segments GNNQQNY and NNQQNY[22–24] showed that short peptides can themselves form both fibrils and closely related microcrystals, the latter capable of revealing atomic structures. From these structures, it was evident how short segments form fibrils: the cross-β spine consists of a pair of β-sheets; each sheet is formed from extended strands of the segment, hydrogen-bonding up and down the sheet to identical molecules, all perpendicular to the axis of the fibril. Two sheets mate tightly at a completely dry interface. At this interface, the residue side chains intermesh with close complementarity, in what we term a steric zipper. The pair-of-sheets motif, with its dry, steric-zipper interface repeated along the entire length of the needle-shaped crystal, accounts for the elongated shape of the crystal and presumably of the fibril. In fibrils formed from full proteins, the extra-spine regions may remain on the periphery of the spine, in some cases in native-like conformation[25,26].

These initial results encouraged us to identify fibril-forming segments in other disease-related and fibril-forming proteins, and to determine structures for these new cross-β spines. In fact, in this class of proteins, we have identified one or more such segments in every protein we examined. Identification of these segments was based on a combination of bioinformatic and experimental procedures[23,27–29], guided in some cases by the published work of others. So far we have identified some 30 such segments from 14 different proteins (see Supplementary Table 1). Most of these fibril-forming segments also form needle-shaped microcrystals, varying in length but rarely larger than 2×2 μm in cross section (Fig. 1). Several control segments, predicted not to form fibrils, did not form fibrils[29]. The wide variety of protein sequences that form fibrils ranges from highly polar (for example, NNQQ) to highly apolar (for example, MVGGVV), and from small side chains (for example, SSTSAA) to large (for example, FYLLYY). Despite their variation, these sequences have a property in common: their self-complementary binding.

Fibrils related to microcrystals

Although the 13 segment structures are known with high accuracy (resolutions between 0.85 and 2.0 Å and *R*-factors between 0.07 and 0.24), there is lingering uncertainty as to how reflective these structures are of amyloid fibrils. Three types of evidence suggest that the crystal structures of segments we report here are related to structures of the fibrils formed by the same segment, and also to fibrils formed from the entire proteins from which the segments are taken.

[1]Howard Hughes Medical Institute, UCLA-DOE Institute of Genomics and Proteomics, Los Angeles, California 90095-1570, USA. [2]Centre for Crystallographic Studies, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 KBH, Denmark. [3]European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble Cedex, France.

Structural similarities of microcrystals to fibrils. Microcrystals and fibrils often grow under the same conditions, and are sometimes found together in solution. Some fibrils (Fig. 1b, c) appear to grow from tips of microcrystals. In all of the segment crystals, the segment fibrils, and the fibrils of whole proteins that we have tested, β-strands are the principal diffracting feature. The strands extend perpendicular to the long axis of the crystal and of the fibrils. Also, the inter-sheet spacings and interstrand spacings of our crystals are in general accord with the spacings found in amyloid fibrils by X-ray fibre diffraction[7,30]. We show this systematically for microcrystals of segments from Sup35 and amyloid-β protein in Supplementary Fig. 1. This figure compares X-ray fibril diffraction patterns of the full proteins with simulated fibril patterns from our microcrystals, produced by cylindrical averaging of the single microcrystal X-ray data. The comparison shows a good fit of the principal diffraction features, suggesting that the principal structural features of the protein fibrils are closely similar to those within the microcrystals.

Nucleation experiments. Nucleated growth is one of the hallmarks of amyloid fibrils[31], and amyloid nucleation requires equivalence of the molecular structures of the nucleus and the fibril[32]. To study nucleated growth, we prepared crystalline seeds from the insulin segment LVEALYL, from which we have determined the steric-zipper structure of VEALYL. These seeds shorten the characteristic lag time for the growth of insulin fibrils as much as seeds prepared from fibrils of the entire insulin molecule (Supplementary Fig. 2), whereas a non-fibril-forming segment of insulin fails to shorten the lag time. This is strong evidence for the involvement of LVEALYL in insulin fibrils. We also found that full-length Sup35 accelerates GNNQQNY peptide aggregation[23]. Both studies suggest a structural similarity between peptide segments and fibrils of their parent proteins.

Mutations in fibril-forming segments affect fibril formation of whole proteins. If the segments of our structures are in fact in the amyloid spines of their parent proteins, we would expect that mutations in the segments would diminish fibril formation, and that these segments would show enhanced protection from proton exchange. In fact, as detailed in Supplementary Table 2, work of others has established that mutants within eight of these segments either slow or diminish fibril formation. Other studies have found that three of our segments lie in regions of fibrils that show protection to proton exchange.

In summary, diffraction patterns show that the principal diffracting features of protein fibrils and the corresponding microcrystals are closely similar; the ability of the protein segments to seed fibrils of their parent proteins shows that these segments are involved in protein fibril formation and probably have similar atomic arrangements in microcrystals and protein fibrils; and the fact that mutant sequences of the microcrystalline segments affect fibril formation of their parent proteins suggests that these segments are involved in protein fibril formation. Although none of these experiments proves that the structure of the microcrystal is the same as that of their corresponding protein fibrils, the evidence indicates a structural similarity.

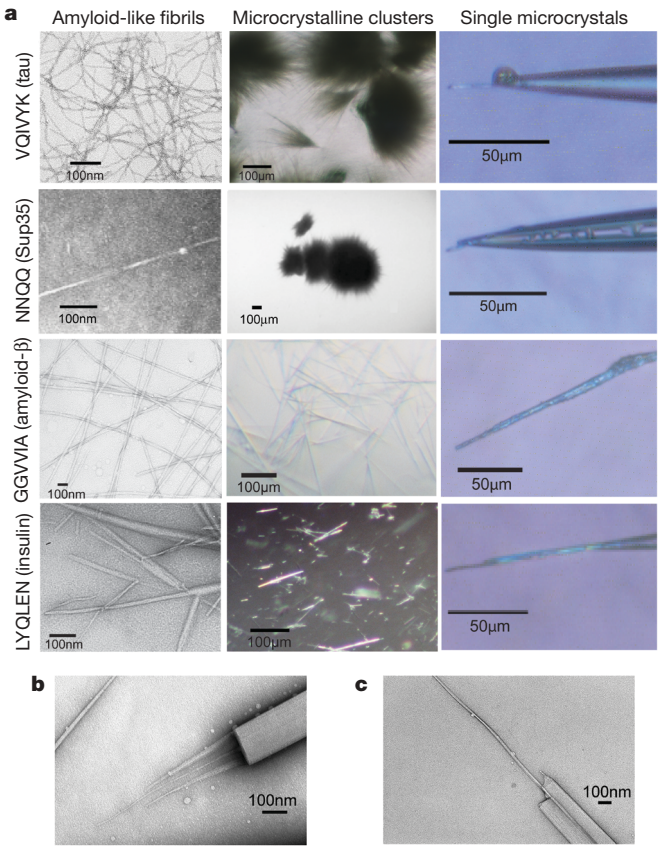


Figure 1 | Amyloid fibrils and microcrystals. a, Electron micrographs of representative amyloid-like fibrils (left), and magnified images of microcrystalline clusters (middle) and single microcrystals mounted for X-ray diffraction (right) of four segments identified from fibril-forming proteins: VQIVYK from tau, NNQQ from Sup35, GGVVIA from amyloid-β, and LYQLEN from insulin. **b**, **c**, Microcrystals of fibril-forming segments of amyloid-forming proteins, appearing to have fibrils growing from their tips. Shown are negatively stained transmission electron micrographs of a microcrystal of segment NFLVHSS from IAPP (**b**), and of VEALYL from insulin (**c**). In both panels, the microcrystals are on the right.

Eight classes of steric zippers

Although varied in sequence, all of our 11 new high-resolution microcrystal structures (Figs 2, 3; Supplementary Table 3; and Supplementary Fig. 3) resemble those of GNNQQNY and NNQQNY[22] in containing extended protein strands that are perpendicular to the needle axes and organized into standard Pauling–Corey β-sheets. Because every one of the structures is built around the steric zipper that we found previously in GNNQQNY and NNQQNY[22], the structures suggest that dry, steric-zipper interfaces between β-sheets are a general principle of protein complementation in amyloid structures. Other examples of extended protein or peptide chains forming such steric-zipper interfaces between protein chains are essentially absent from the PDB (Protein Data Bank) and are rare in the CSD (Cambridge Structural Database), supporting the idea that these interfaces are the defining molecular property of the amyloid state.

Despite their fundamental similarity, the reported structures display variations of the basic steric-zipper structure and thereby expand our understanding of amyloid structure. The structures to date fall into five classes (Figs 2, 3), distinguished by (1) whether their sheets (that is, the strands in their sheets) are parallel or antiparallel, (2) whether sheets pack with the same ('face-to-face') or different ('face-to-back') surfaces adjacent to one another, and (3) whether the sheets are oriented parallel ('up–up') or antiparallel ('up–down') with respect to one another. To distinguish this third type of orientation from the first, we refer to the relative sheet orientations in terms of a given sheet edge facing 'up' or 'down.' Combinations of these three structural arrangements give eight theoretically possible classes of steric zippers, shown in Fig. 4. Examples of classes 1, 2, 4, 7 and 8 are represented in our 13 microcrystal structures (Figs 2, 3).

The steric zippers of class 1 share a basic unit of two parallel, in-register β-sheets with their same sides facing each other ('face-to-face') and both sheet edges facing 'up'. Each sheet is related to its mate by a 2_1 axis parallel to the needle axis of the crystal, which rotates a sheet by 180° about the axis and moves it along the axis by half the spacing between two β-strands (Fig. 3, far left). This symmetry allows the side chains of one sheet to nestle between layers of side chains on the mating sheet, forming the steric zipper. In class 2, the symmetry is a simple translation of one sheet onto its neighbour, resulting in a

'face-to-back, up–up' packing. SNQNNF, from the human prion protein, is thus far the only example of a class 2 zipper. In class 4, like class 2, the apposed, interdigitating faces of the sheets differ from each other ('face-to-back'). However, neighbouring sheets of class 4 are oriented with one sheet's edge facing 'up' and its neighbours' edges 'down'. GGVVIA, from the carboxy terminus of amyloid-β, adopts this orientation. The sheets of classes 7 and 8, like those of the 'parallel' classes 1–4, contain β-strands in register, but within each sheet, adjacent strands run in opposite directions. Antiparallel β-sheets in amyloid-like fibrils have been anticipated from previous

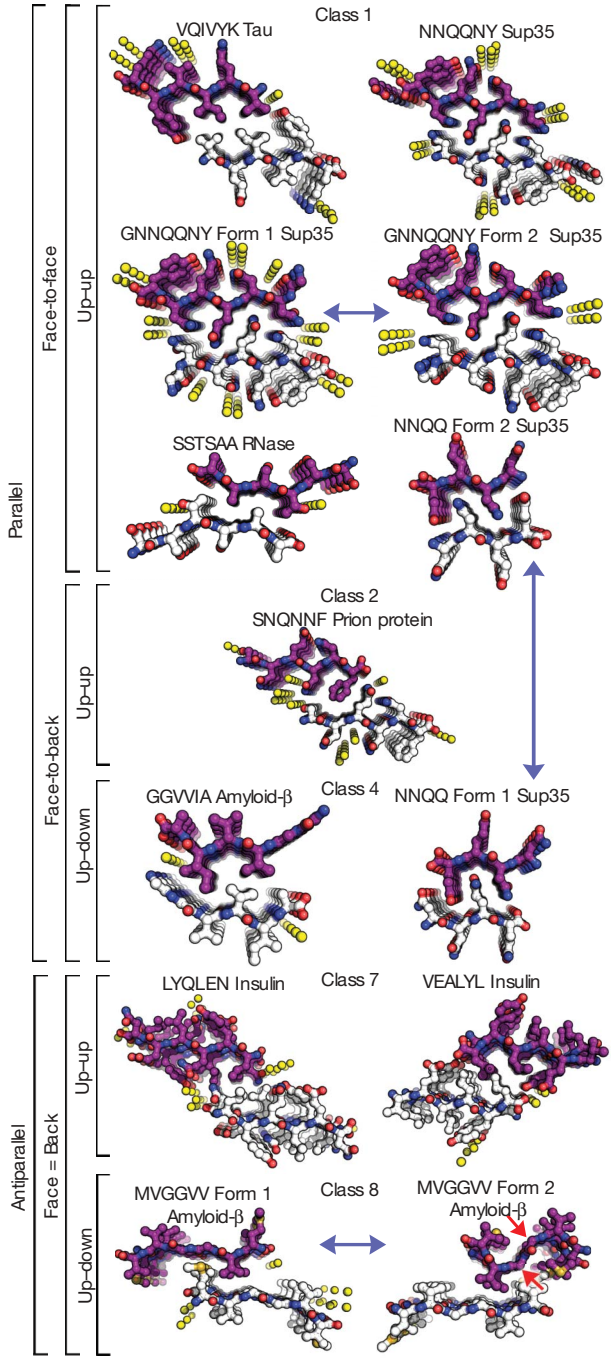


Figure 2 | Thirteen atomic-resolution structures for peptide segments of fibril-forming proteins. See text for details of nomenclature. A two-sheet motif of each structure is depicted in projection down the needle crystal axis, showing only the top members of ~10^5 stacked segments in each crystalline sheet. A dry, steric-zipper interaction is evidenced by the interdigitation of side chains between sheets. Carbon atoms are shown as purple or white, nitrogen as blue, and oxygen as red. Water molecules are shown as yellow spheres. NNQQNY also contains zinc acetate. Zippers are grouped by class (1, 2, 4, 7, 8); see text for details. Previously reported Sup35 zippers[22] belong to class 1. The three pairs of structures related by blue double-headed arrows are polymorphic pairs (forms 1 and 2; see text for details). The red arrows point to the 90° bend in the upper sheet of MVGGVV form 2.

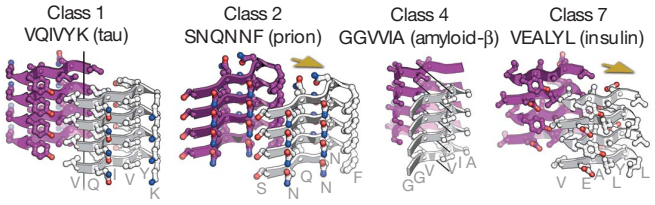


Figure 3 | 3D views of representative steric zipper structures of classes 1, 2, 4 and 7, showing the front sheet in silver and the rear sheet in purple. Oxygen atoms are red; nitrogen atoms are blue. Black lines show crystallographic 2_1 symmetry axes, and the yellow arrows show translational symmetry. The value of the shape complementarity parameter[46], S_C, for GGVVIA ($S_C = 0.92$) is the largest value we have found for any protein interface, consistent with the higher toxicity and lower solubility of amyloid-β(1−42) than (1–40).

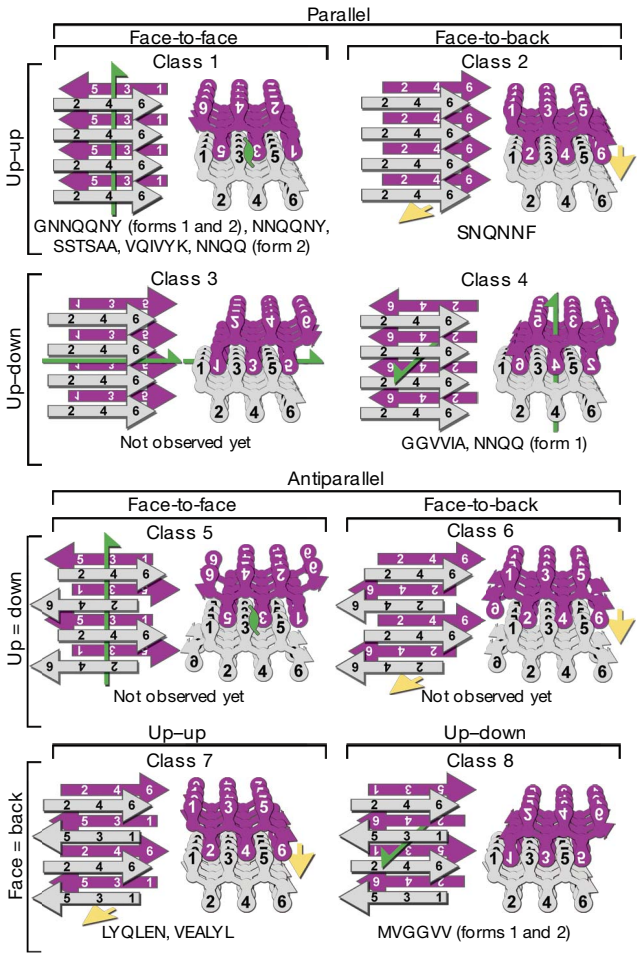


Figure 4 | The eight classes of steric zippers. Two identical sheets can be classified by: the orientation of their faces (either 'face-to-face' or 'face-to-back'), the orientation of their strands (with both sheets having the same edge of the strand 'up', or one 'up' and the other 'down'), and whether the strands within the sheets are parallel or antiparallel. Both side views (left) and top views (right) show which of the six residues of the segment point into the zipper and which point outward. Green arrows show two-fold screw axes, and yellow arrows show translational symmetry. Below each class are listed protein segments that belong to that class.

X-ray diffraction studies[33–36]. 'Face-to-back' and 'Face=back' arrangements (classes 2, 4, 6 and 8) lead naturally to adhesion of further β-sheets, favouring macroscopic tubes and sheets (for example, LYQLEN, Fig. 1a), as well as the fibrils usually seen with the pair-of-sheets motif of class 1 (Fig. 1a). To date, we have not encountered microcrystals with steric zippers of classes 3, 5 or 6.

Multiple segments and prion strains

In identifying fibril-forming segments within known fibril-forming proteins (Supplementary Table 1), we found that several proteins contain more than one fibril-forming segment. Examples are Sup35, tau, amyloid-β, β_2-microglobulin, insulin, IAPP, α-synuclein and PrP. To date, we have determined structures of different segments of both insulin and amyloid-β, finding significantly different structures for segments from the same protein. For example, GGVVIA from amyloid-β forms parallel sheets, whereas the overlapping segment MVGGVV forms antiparallel sheets. Thus fibrils formed from entire proteins could conceivably contain more than a single type of paired sheets, requiring more elaborate models (see, for example, refs 37, 38). Alternatively, protein fibrils may contain sheets built from more than a single type of protein segment. Yet another possibility is polymorphic fibrils of the same protein, each with its own steric-zipper structure. Full understanding of the organization of such fibrils will require structural studies of these more complex sheet structures.

Three of the segments shown in Fig. 2 each form two polymorphs, offering a glimpse of the possible molecular basis of the phenomena of prion strains and amyloid polymorphism[39–45]. In prion strains, a single protein sequence encodes different phenotypes, attributed recently to 'distinguishable, self-propagating structural features'[44] or different 'prion conformations'[45]. A possible molecular explanation for these effects is alternative steric zippers formed by the same sequence segment. For example, in Fig. 2, there are two polymorphic structures of NNQQ. In form 2 (class 1), the sheets are 'face-to-face', whereas they are 'face-to-back' in form 1 (class 4). We suggest that fibrils built on these two arrangements would be distinctly different in structure, and probably different in properties. Notice that the two polymorphs of GNNQQNY show essentially identical steric zippers, but the packing of the steric zippers within their crystals (Supplementary Fig. 3) differ. This suggests that crystal-packing forces do not substantially distort the basic structure of the steric zipper, although it is likely that the flatness of the sheets is influenced by packing forces.

In several of these crystal structures (as shown in Supplementary Fig. 3), sheet-to-sheet interactions involve more than just one zipper, which offers further possibilities for polymorphic structures. This is in contrast to the structures of GNNQQNY form 1, NNQQNY and VQIVYK, in which a single type of steric-zipper interaction accounts for the majority of the buried surface area between sheets (Supplementary Fig. 3). In the cases of MVGGVV form 1, LYQLEN and SSTSAA, the amount of buried surface area on a given β-sheet face is split roughly equally between two interfaces. In the example of NNQQ form 2, there are two different 'face-to-face' steric zippers. Such structures show multiple choices for tight packing of two sheets together, suggesting that sheets of the full-length protein could pack together in multiple ways to form polymorphic fibrils.

Discussion

When the 30 fibril-forming segments given in Supplementary Table 1 and the 13 microcrystal structures in Fig. 2 are considered together, they reinforce the hypothesis that microcrystal structures reveal fundamental structural features of amyloid fibrils. First, within each fibril-forming protein that we have examined, we can identify at least one segment of 4–12 residues which itself forms fibrils in isolation from the rest of the protein chain. That every fibril-forming protein contains a fibril-forming segment suggests that this segment may drive fibril formation of the protein. Second, as related above, the fibril-forming peptide segments are linked to fibrils of their parent proteins by structural, nucleation or mutational data, suggesting that the crystal structures reveal essential features of fibril structure, although there is no definitive proof that the atomic interactions are identical.

On the basis of the crystal structures reported here, as well as of earlier work by others, we summarize our observations of amyloid fibrils. (1) A 4–7-residue segment of sequence is sufficient to form a fibril in at least a dozen disease-related proteins. (2) The fundamental unit of amyloid-like fibrils is a steric zipper, formed by two tightly interdigitated β-sheets, with the possibility of more complicated geometries with multiple steric zippers. That is, the ~30 microcrystalline, fibril-forming segments of Supplementary Table 1 and the dry steric-zipper structures of Fig. 2 together suggest that amyloid diseases are similar not only on the fibril level, but also on the molecular level. (3) On the basis of the discovery of steric zippers in several disease-associated proteins, it seems likely that the process of fibril formation starts by the unmasking of zipper-forming segments in several identical protein molecules, permitting them to stack into β-sheets and the sheets to interdigitate. Recruitment of monomers into pre-formed fibrils is expected to be more rapid than nucleation, because recruitment requires only one molecule at a time to unmask its fibril-forming sequence, but formation of the steric-zipper nucleus requires several molecules to unmask their zipper-forming segments simultaneously. That is, the common feature of all these structures—the dry steric zipper—is compatible with slow fibril nucleation and faster fibril growth, the commonly observed kinetic characteristics of fibril formation[31]. (4) The finding of distinct crystalline polymorphs of the same zipper-forming segments offers a molecular-level hypothesis for amyloid polymorphism and prion strains, which awaits verification or disproof. (5) There are potentially eight classes of steric zippers, five of which have been experimentally confirmed.

METHODS

A full description of methods is given in Supplementary Information. Briefly, fibrils and microcrystals of the peptides were grown through the dissolution of lyophilized, synthetic peptide in water or buffers. Crystals were generally grown by the hanging drop method using standard crystallization screens. X-ray data were collected at ESRF beamline ID13, processed and scaled using standard software, and phased by molecular replacement with idealized β-strands. Thioflavin T fluorescence at 482 nm wavelength was used to monitor insulin fibril formation. Equal volumes of the various seeds were added to identical reactions, and each experiment was recorded for six replicates.

Received 30 November 2006; accepted 19 February 2007.
Published online 29 April 2007.

1. Westermark, P. Aspects on human amyloid forms and their fibril polypeptides. *FEBS J.* **272,** 5942–5949 (2005).
2. Westermark, P. *et al.* Amyloid: toward terminology clarification. Report from the Nomenclature Committee of the International Society of Amyloidosis. *Amyloid* **12,** 1–4 (2005).
3. Cohen, A. S. & Calkins, E. Electron microscopic observations on a fibrous component in amyloid of diverse origins. *Nature* **183,** 1202–1203 (1959).
4. Rochet, J. C., Lansbury, P. T. Jr., Amyloid fibrillogenesis: themes and variations. *Curr. Opin. Struct. Biol.* **10,** 60–68 (2000).
5. Astbury, W. T., Dickinson, S. & Bailey, K. The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem. J.* **29,** 2351–2360 (1935).
6. Geddes, A. J., Parker, K. D., Atkins, E. D. & Beighton, E. "Cross-β" conformation in proteins. *J. Mol. Biol.* **32,** 343–358 (1968).
7. Sunde, M. & Blake, C. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50,** 123–159 (1997).
8. Eanes, E. D. & Glenner, G. G. X-ray diffraction studies on amyloid filaments. *J. Histochem. Cytochem.* **16,** 673–677 (1968).
9. Sunde, M. *et al.* Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **273,** 729–739 (1997).
10. Ritter, C. *et al.* Correlation of structural elements and infectivity of the HET-s prion. *Nature* **435,** 844–848 (2005).
11. Sikorski, P. & Atkins, E. New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils. *Biomacromolecules* **6,** 425–432 (2005).
12. Petkova, A. T. *et al.* A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl Acad. Sci. USA* **99,** 16742–16747 (2002).

13. Jaroniec, C. P. *et al.* High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **101,** 711–716 (2004).
14. Krishnan, R. & Lindquist, S. L. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* **435,** 765–772 (2005).
15. Makin, O. S. & Serpell, L. C. Structures for amyloid fibrils. *FEBS J.* **272,** 5950–5961 (2005).
16. Lührs, T. *et al.* 3D structure of Alzheimer's amyloid-β(1–42) fibrils. *Proc. Natl Acad. Sci. USA* **102,** 17342–17347 (2005).
17. Török, M. *et al.* Structural and dynamic features of Alzheimer's Aβ peptide in amyloid fibrils studied by site-directed spin labeling. *J. Biol. Chem.* **277,** 40810–40815 (2002).
18. Paravastu, A. K., Petkova, A. T. & Tycko, R. Polymorphic fibril formation by residues 10–40 of the Alzheimer's β-amyloid peptide. *Biophys. J.* **90,** 4618–4629 (2006).
19. Williams, A. D. *et al.* Mapping Aβ amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol. Biol.* **335,** 833–842 (2004).
20. Kheterpal, I., Zhou, S., Cook, K. D. & Wetzel, R. Aβ amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *Proc. Natl Acad. Sci. USA* **97,** 13597–13601 (2000).
21. Ferguson, N. *et al.* General structural motifs of amyloid protofilaments. *Proc. Natl Acad. Sci. USA* **103,** 16248–16253 (2006).
22. Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435,** 773–778 (2005).
23. Balbirnie, M., Grothe, R. & Eisenberg, D. S. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated β-sheet structure for amyloid. *Proc. Natl Acad. Sci. USA* **98,** 2375–2380 (2001).
24. Diaz-Avalos, R. *et al.* Cross-β order and diversity in nanocrystals of an amyloid-forming peptide. *J. Mol. Biol.* **330,** 1165–1175 (2003).
25. Baxa, U. *et al.* Architecture of Ure2p prion filaments: the N-terminal domains form a central core fiber. *J. Biol. Chem.* **278,** 43717–43727 (2003).
26. Sambashivan, S., Liu, Y., Sawaya, M. R., Gingery, M. & Eisenberg, D. Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature* **437,** 266–269 (2005).
27. Thompson, M. J. *et al.* The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl Acad. Sci. USA* **103,** 4074–4078 (2006).
28. Ivanova, M. I., Sawaya, M. R., Gingery, M., Attinger, A. & Eisenberg, D. An amyloid-forming segment of β2-microglobulin suggests a molecular model for the fibril. *Proc. Natl Acad. Sci. USA* **101,** 10584–10589 (2004).
29. Ivanova, M. I., Thompson, M. J. & Eisenberg, D. A systematic screen of β_2-microglobulin and insulin for amyloid-like segments. *Proc. Natl Acad. Sci. USA* **103,** 4079–4082 (2006).
30. Fändrich, M. & Dobson, C. M. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21,** 5682–5690 (2002).
31. Harper, J. D., Lansbury, P. T. Jr., Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu. Rev. Biochem.* **66,** 385–407 (1997).
32. Jarrett, J. T., Lansbury, P. T. Jr., Amyloid fibril formation requires a chemically discriminating nucleation event: studies of an amyloidogenic sequence from the bacterial protein OsmB. *Biochemistry* **31,** 12345–12352 (1992).
33. Sikorski, P., Atkins, E. D. & Serpell, L. C. Structure and texture of fibrous crystals formed by Alzheimer's Aβ(11–25) peptide fragment. *Structure (Camb.)* **11,** 915–926 (2003).
34. Makin, O. S., Atkins, E., Sikorski, P., Johansson, J. & Serpell, L. C. Molecular basis for amyloid fibril formation and stability. *Proc. Natl Acad. Sci. USA* **102,** 315–320 (2005).
35. Halverson, K., Fraser, P. E., Kirschner, D. A., Lansbury, P. T. Jr., Molecular determinants of amyloid deposition in Alzheimer's disease: conformational studies of synthetic β-protein fragments. *Biochemistry* **29,** 2639–2644 (1990).
36. Serpell, L. C. & Smith, J. M. Direct visualisation of the β-sheet structure of synthetic Alzheimer's amyloid. *J. Mol. Biol.* **299,** 225–231 (2000).
37. Kajava, A. V., Baxa, U., Wickner, R. B. & Steven, A. C. A model for Ure2p prion filaments and other amyloids: the parallel superpleated β-structure. *Proc. Natl Acad. Sci. USA* **101,** 7885–7890 (2004).
38. Kajava, A. V., Aebi, U. & Steven, A. C. The parallel superpleated β-structure as a model for amyloid fibrils of human amylin. *J. Mol. Biol.* **348,** 247–252 (2005).
39. King, C. Y. & Diaz-Avalos, R. Protein-only transmission of three yeast prion strains. *Nature* **428,** 319–323 (2004).
40. Tanaka, M., Chien, P., Naber, N., Cooke, R. & Weissman, J. S. Conformational variations in an infectious protein determine prion strain differences. *Nature* **428,** 323–328 (2004).
41. Petkova, A. T. *et al.* Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science* **307,** 262–265 (2005).
42. Chien, P. & Weissman, J. S. Conformational diversity in a yeast prion dictates its seeding specificity. *Nature* **410,** 223–227 (2001).
43. Jones, E. M. & Surewicz, W. K. Fibril conformation as the basis of species- and strain-dependent seeding specificity of mammalian prion amyloids. *Cell* **121,** 63–72 (2005).
44. Diaz-Avalos, R., King, C. Y., Wall, J., Simon, M. & Caspar, D. L. Strain-specific morphologies of yeast prion amyloid fibrils. *Proc. Natl Acad. Sci. USA* **102,** 10165–10170 (2005).
45. Tanaka, M., Collins, S. R., Toyama, B. H. & Weissman, J. S. The physical basis of how prion conformations determine strain phenotypes. *Nature* **442,** 585–589 (2006).
46. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–950 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. L. D. Caspar, D. Anderson, D. Cascio, M. Gingery, M. Graf and K. Wüthrich for discussions, and the NSF, the NIH and the HHMI for support. S.A.S. was supported by an NSF IGERT training grant and M.I.A. by an NIH National Research Service Award.

Author Contributions M.R.S., S.S., R.N. and M.I.I contributed equally to this work.

Author Information The 11 new structures shown in Fig. 2, and their structure factors, have been deposited in the Protein Data Bank with accession codes as follows: GNNQQNY form 2, 2OMM; NNQQ form 1, 2ONX; NNQQ form 2, 2OLX; VEALYL, 2OMQ; LYQLEN, 2OMP; VQIVYK, 2ON9; GGVVIA, 2ONV; MVGGVV form 1, 2ONA; MVGGVV form 2, 2OKZ; SSTSAA, 2ONW; SNQNNF, 2OL9. In addition, at http://www.doe-mbi.ucla.edu/~sawaya/chime/xtalpept/ we offer, for each microcrystal structure, coordinates of the asymmetric unit, the unit cell, many unit cells, and a pair of sheets. This site also offers Chime pages that illustrate the steric zipper, polar zippers, crystal packing and water exclusion. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.E. (david@mbi.ucla.edu).

LETTERS

An unusually brilliant transient in the galaxy M85

S. R. Kulkarni[1], E. O. Ofek[1], A. Rau[1], S. B. Cenko[2], A. M. Soderberg[1], D. B. Fox[3], A. Gal-Yam[1], P. L. Capak[1], D. S. Moon[2], W. Li[4], A. V. Filippenko[4], E. Egami[5], J. Kartaltepe[6] & D. B. Sanders[6]

Historically, variable and transient sources have both surprised astronomers and provided new views of the heavens. Here we report the discovery of an optical transient in the outskirts of the lenticular galaxy Messier 85 in the Virgo cluster. With a peak absolute *R* magnitude of −12, this event is distinctly brighter than novae, but fainter than type Ia supernovae (which are expected in a population of old stars in lenticular galaxies). Archival images of the field do not show a luminous star at that position with an upper limit in the *g* filter of about −4.1 mag, so it is unlikely to be a giant eruption from a luminous blue variable star. Over a two-month period, the transient source emitted radiation energy of almost 10^{47} erg and subsequently faded in the optical sky. It is similar to, but six times more luminous at peak than, an enigmatic transient in the galaxy M31 (ref. 1). A possible origin of M85 OT2006-1 is a stellar merger. If so, searches for similar events in nearby galaxies will not only allow study of the physics of hyper-Eddington sources, but also probe an important phase in the evolution of stellar binary systems.

On January 7, 2006, the Lick Observatory Supernova Search team, in their daily circular, reported a source with apparent unfiltered magnitude of ~19.3 projected 2.3 kpc from the centre of the lenticular (Hubble type S0) galaxy Messier 85 (M85; also known as NGC 4382) a member of the Virgo cluster of galaxies; see Fig. 1. We believe (see below) that the transient lies within M85 and thus we name the source the M85 Optical Transient 2006-1 (OT2006-1). Fortuitously, this field was observed by the Hubble Space Telescope three years earlier. From the archival Hubble Space Telescope data we derive a pre-explosion limiting magnitude of F475W = −4.3.

On 8 January 2006 UT we initiated an optical photometric campaign with the automated Palomar 60-inch telescope (see Fig. 2 and Supplementary Table). The light curve, with its plateau of 70 days, is unlike that of a type Ia supernova. The plateau duration is also too short for an outburst from a luminous blue variable (LBV, such as η Carina). We began a programme of spectroscopic observations with the Palomar Hale and Keck I telescopes (Fig. 3). The Palomar spectrum obtained on 8 January 2006 UT did not contain any strong emission feature; the spectrum could be adequately described by a blackbody spectrum with effective temperature of approximately $T_{\rm eff} \approx 4{,}600$ K. Likewise, the 3 February 2006 UT Keck spectrum was also featureless but unfortunately did not cover Hα. The 23

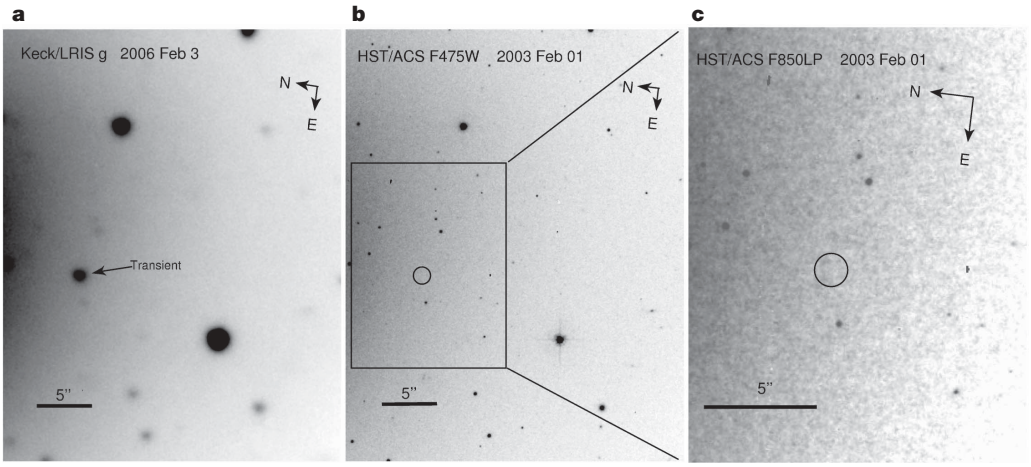


Figure 1 | Optical images of the field around M85 OT2006-1 obtained at two epochs. Data were obtained with the Low-Resolution Imager and Spectrograph (LRIS)[11] at Keck (**a**) on 3 February 2006, and the Advanced Camera for Surveys aboard the Hubble Space Telescope (**b**, with a F475W filter; **c**, with a F850LP filter) on 1 February 2003. The event is located about 30″ from the centre of M85 at $\alpha = 12\,\mathrm{h}\,25\,\mathrm{min}\,23.82\,\mathrm{s}$ and $\delta = 18^\circ\,10'\,56.2''$ (J2000). After registering the Keck image to the Hubble Space Telescope image (the root mean square of the transformation was 40 mas) we were able to establish the following limits for a precursor object (progenitor star): 26.8 mag in the F475W filter (exposure 750 s) and 24.7 mag in the F850LP filter (exposure time 1,150 s). These limits exclude an LBV[2] origin (for which $M_V \approx -8$ mag). Furthermore, we find no evidence for young stars (supergiants, clusters and H II regions). An analysis of Spitzer Space Telescope Infared Array Camera data obtained on 21 December 2004 resulted in 3σ upper limits of 25, 30, 60 and 75 μJy at 3.6, 4.5, 5.8 and 8.0 μm, respectively. The Lick Observatory Supernova Search observed M85 two hundred and twenty times from 2000 to 2006. We found no transient at the position of M85 OT2006-1 to (roughly *R*-band) magnitudes ranging from 20 to 21. No X-ray emission was detected in a Chandra X-ray Observatory observation[12] obtained in June 2002 with a flux upper limit of 2.7×10^{-4} counts s^{-1} in the 0.3–10-keV band. Circles indicate the position of the transient in **b** and **c**.

[1]Caltech Optical Observatories 105-24, [2]Space Radiation Laboratory 220-47, California Institute of Technology, California 91125, USA. [3]Department of Astronomy, Pennsylvania State University, State College, Pennsylvania 16802, USA. [4]Astronomy Department, 601 Campbell Hall, University of California, Berkeley, California 94720, USA. [5]Steward Observatory, 933 N. Cherry Avenue, University of Arizona, Tucson, Arizona 85721, USA. [6]Institute of Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA.

and 24 February 2006 UT Keck spectra showed a similar continuum but a number of emission lines were readily detected (Fig. 3). These latter spectra were the deepest, so it is likely that the lines were seen as a result of better sensitivity.

We associate the strongest lines at wavelengths $\lambda \simeq 6{,}587$ Å and $\lambda \simeq 4{,}874$ Å with Hα and Hβ, respectively. Assuming this identification is correct, the mean heliocentric (peak) velocity of the pair is $880 \pm 130\,\mathrm{km\,s^{-1}}$. We were unable to conclusively identify the remaining lines but do note that the spectra of many hypergiants contain a number of unidentified[2,3] emission lines.

The systemic velocity[4] of M85 is $729 \pm 2\,\mathrm{km\,s^{-1}}$ and the velocity dispersion[5] in the vicinity of the optical transient is $200\,\mathrm{km\,s^{-1}}$. The peak velocity of the Balmer lines is thus consistent with M85 OT2006-1 being located in M85. So if M85 is the host galaxy (for which we adopt a distance of 15 Mpc, the standard distance to the Virgo cluster[6]) the absolute R-band magnitude of M85 OT2006-1 is -12 mag. This peak flux is ten times brighter than the brightest nova but (at least) ten times less luminous than type Ia supernovae (the sort expected in a lenticular galaxy). The narrow linewidth of the Hα line, $\sim 350 \pm 140\,\mathrm{km\,s^{-1}}$ (see Fig. 3), argues independently against an origin of either a nova or a supernova (including type II).

The Galactic foreground extinction towards M85 is negligible, $A_R = 0.08$ (ref. 7). The source intrinsic attenuation can be derived by comparing the observed ratio of the emission lines fluxes of Hα ($(3.2 \pm 0.2) \times 10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$) and H$\beta$ ($(0.9 \pm 0.1) \times 10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$) and the theoretical value of 3.05 (case B recombination[8], low-density limit, $T = 5{,}000$ K). We estimate $E(B-V) = 0.14^{+0.17}_{-0.14}$, which corresponds to an R-band extinction of $0.40^{+0.48}_{-0.40}$ mag. This is too low to explain the unusual colour and temperature of M85 OT2006-1 with a strongly absorbed nova, supernova or LBV.

We searched archival data from the Hubble Space Telescope, the Spitzer Space Telescope and the Chandra X-ray Observatory in an attempt to constrain the progenitor. There is no evidence for a bright progenitor, and nor do we see tracers of massive star progenitors (see Fig. 1). This finding (and the shorter duration) rule out that M85 OT2006-1 is an LBV because LBVs are among[2] the brightest stars, $M_V < -8$. We note that M85 is a galaxy composed of old stars with a possible trace of a spiral arm. We conclude that the M85 OT2006-1 probably arises from a population of stars with masses of a few times the mass of the Sun ($M_\odot$) or less.

We now turn to the physical parameters of M85 OT2006-1. The bolometric luminosity flux (as traced by $4\pi d^2 \nu f_\nu$, where f_ν is the spectral flux density at frequency ν) of M85 OT2006-1 peaks at

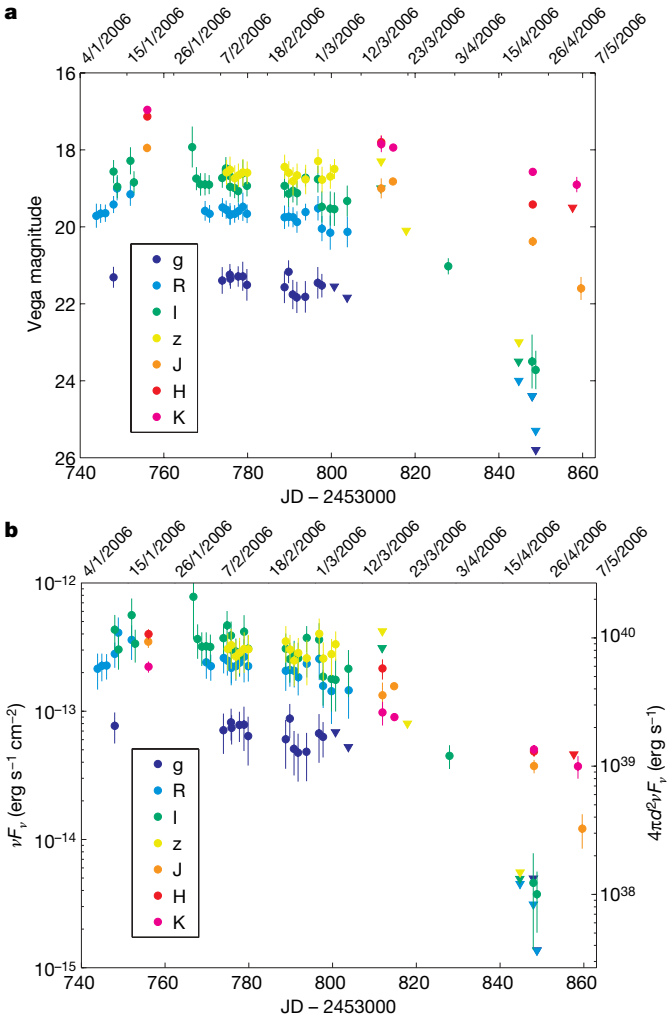


Figure 2 | Temporal evolution of M85 OT2006-1. **a**, Observed light curve uncorrected for Galactic foreground extinction. **b**, νf_ν including foreground extinction correction of $A_V = 0.08$ (ref. 7). Data for the plots are given in the Supplementary Table and come from the following sources (bands given in parentheses): Palomar 60-inch (P60; *gRIz*), the Large Format Camera (LFC) on the Palomar Hale 200-inch (P200; *zRI*), the Widefield Infrared Camera (WIRC) on P200 (*JHK*), LRIS on the Keck-I 10-m telescope (*gRI*), Persson's Auxilary Nasmyth Infrared Camera (PANIC) on the Magellan 6.5-m Baade telescope (*YJK*), the Near Infared Echelle Spectrograph (NIRSPEC) on the Keck-II 10-m telescope (*JHK*) and the Wide Field Infrared Camera (WFCAM) on the 3.8-m United Kingdom Infrared Telescope (UKIRT; *JHK*). Error bars are 1σ. JD, Julian day.

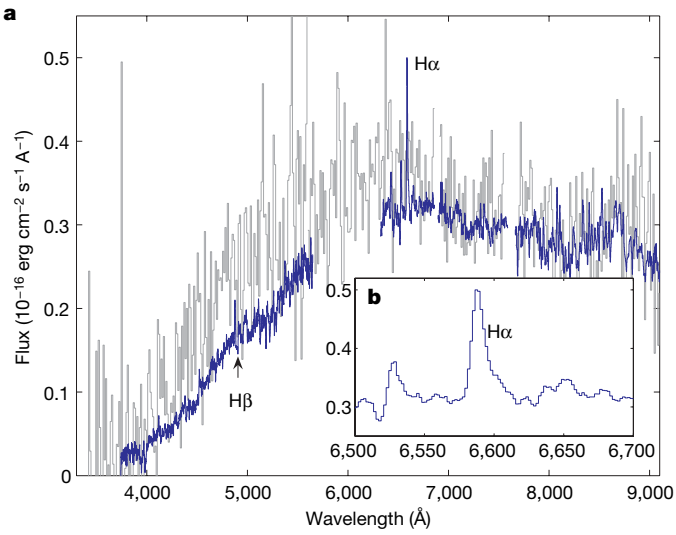


Figure 3 | Optical spectra of M85 OT2006-1. Data were obtained with the Double Beam Spectrograph[13] (DBSP) at the Palomar Hale 200-inch telescope (grey line, 1,800 s integration, 2006 January 8.53 UT) and Keck/LRIS (blue line, 3,000 s, 2006 February 24.59 UT). Neither strong emission nor absorption features are seen in the (native) DBSP spectrum. Specifically we placed a limit of $6 \times 10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ for an emission line in the vicinity of Hα. In the LRIS red channel spectrum the brightest emission feature is at $\lambda = 6{,}587$ Å (flux of $(3.2 \pm 0.2) \times 10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$), which we identify with redshifted Hα. The velocity of the line centre is $1{,}020 \pm 150\,\mathrm{km\,s^{-1}}$ (see inset). On the blue side, the strongest feature is at $\lambda = 4{,}875$ Å corresponding to redshifted ($700 \pm 100\,\mathrm{km\,s^{-1}}$) H$\beta$ (flux of $(0.9 \pm 0.1) \times 10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$). The equivalent widths are 10 ± 1 Å (Hα) and 5 ± 1 Å (Hβ). The full-width at half-maximum (FWHM) of the Hα line, after accounting for the instrumental FWHM, is $350 \pm 140\,\mathrm{km\,s^{-1}}$. In addition, we detect the following emission lines (central wavelengths, fluxes in units of $10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ and typical uncertainties of 1 Å): 4,115 Å (0.3 ± 0.1), 6,428 Å (0.9 ± 0.1), 6,527 Å (1.5 ± 0.4), 8,079 Å (0.8 ± 0.1) and 8,106 Å (0.7 ± 0.1). Further LRIS spectra were obtained on 3 and 23 February 2006 UT (not shown here). The 3 February 2006 LRIS spectrum did not include the Hα wavelength. For this spectrum, using a sliding 10 Å window, we were able to set a 3σ upper limit of $6 \times 10^{-18}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ in the vicinity of Hβ.

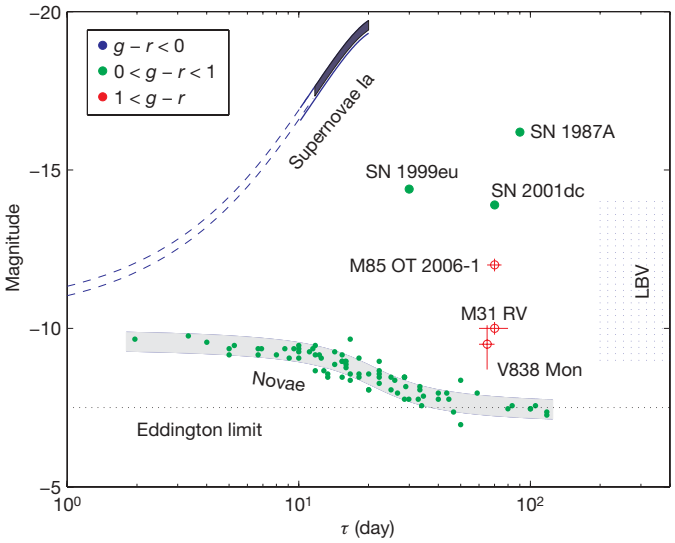


Figure 4 | Phase space of cosmic explosive (supernovae) and eruptive (novae and LBVs) transients. The vertical axis is the peak brightness in the *R* band and the horizontal axis is the duration of the event (τ). Events are represented by circles with the colour at peak magnitude coded as indicated in the key. M85 OT2006-1 and the M31 RV (red variable; ref. 1) (1σ error bars) clearly stand out in this figure in the following respects: they (1) are brighter than novae but (2) are less luminous than most supernovae (especially of type Ia, indicated with a 2σ brightness band) and (3) are distinctly red in colour when compared to subluminous core collapse supernovae (such as SN 1987A). Finally, the two events, unlike LBVs and core collapse supernovae, do not arise in star-forming regions. For any reasonable progenitor mass, both events exhibit hyper-Eddington peak luminosities, similar to the sources V838 Mon[10,14]. Furthermore, both sources are also characterized by low expansion velocity ($<1{,}000\ \mathrm{km\ s^{-1}}$) and a strong redward evolution of the peak frequency. For these objects τ is the 'plateau' timescale. For novae, τ is the timescale in which the nova fades by two magnitudes. Filled circles show the positions of 82 novae observed[15,16] in Messier 31 (assuming[17] $V - R = 0.56$ at peak). The brightest stars in our Galaxy are highly variable, but these objects (marked LBV) are clearly distinguished by long variability timescales and a high quiescent magnitude. The dashed line ($R = -7.5$ mag) is the Eddington limit for a 1 $M_\odot$ G-type object.

$L_{\mathrm{peak}} \approx 2 \times 10^{40}\ \mathrm{erg\ s^{-1}}$. Over the first two months the total radiated energy is about $E_{\mathrm{rad}} \approx 6 \times 10^{46}$ erg. The inferred blackbody radius (in AU) of the object is substantial:

$$R = \left[L_{\mathrm{peak}} / \left(4\pi\sigma_{\mathrm{B}} T_{\mathrm{eff}}^4\right)\right]^{1/2} \approx 17 \left(T_{\mathrm{eff}}/4{,}600\ \mathrm{K}\right)^{-2}$$

where σ_{B} is the Stefan–Boltzmann constant.

The closest analogue to M85 OT2006-1 is M31 RV, a bright event[1] (serendipitously) found in the bulge of Messier 31 and still lacking a satisfactory explanation. The extraordinary brilliance of M85 OT2006-1 (Fig. 4) makes it doubly mysterious. The Galactic transient V838 Monocerotis (ref. 9, while considerably less luminous (Fig. 4), exhibits similar plateau light curves and redward evolution of the broadband spectrum.

The distinctive physical parameters (relative to novae and supernovae; see Fig. 4) and the potential connection to a fundamental stellar process (merger)[10] may warrant coining a name. We suggest the simple name 'luminous red nova', in which the adjectives highlight the principal characteristics of M85 OT2006-1. Statistics (especially including the nature of the host galaxies) and follow-up studies should help astronomers to unravel the origin of these enigmatic transients and also the physics of hyper-Eddington sources.

Received 2 June 2006; accepted 30 March 2007.

1. Rich, R. M., Mould, J., Picard, A., Frogel, J. A. & Davies, R. Luminous M giants in the bulge of M31. *Astrophys. J.* **341,** L51–L54 (1989).
2. Humphreys, R. M. & Davidson, K. The luminous blue variables: Astrophysical geysers. *Publ. Astron. Soc. Pacif.* **106,** 1025–1051 (1994).
3. Humphreys, R. M. *et al.* M33's variable A: a hypergiant star more than 35 years in eruption. *Astron. J.* **131,** 2105–2113 (2006).
4. Smith, R. J., Lucey, J. R., Hudson, M. J., Schlegel, D. J. & Davies, R. L. Streaming motions of galaxy clusters within 12000 km s^{-1}—I. New spectroscopic data. *Mon. Not. R. Astron. Soc.* **313,** 469–490 (2000).
5. Emsellem, E. *et al.* The SAURON project—III. Integral-field absorption-line kinematics of 48 elliptical and lenticular galaxies. *Mon. Not. R. Astron. Soc.* **352,** 721–743 (2004).
6. Freedman, W. L. *et al.* Final results from the Hubble Space Telescope Key Project to measure the Hubble constant. *Astrophys. J.* **553,** 47–72 (2001).
7. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500,** 525–553 (1998).
8. Osterbrock, D. E. *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* Ch. 4 (University Science Books, Mill Valley, California, 1989).
9. Brown, N. J. *et al.* Peculiar variable in Monoceros. *IAU Circ.* **7785** (2002).
10. Tylenda, R. Evolution of V838 Monocerotis during and after the 2002 eruption. *Astron. Astrophys.* **436,** 1009–1020 (2005).
11. Oke, J. B. *et al.* The Keck low-resolution imaging spectrometer. *Publ. Astron. Soc. Pacif.* **107,** 375–385 (1995).
12. Sivakoff, G. R., Sarazin, C. L. & Irwin, J. A. Chandra observations of low-mass x-ray binaries and diffuse gas in the early-type galaxies NGC 4365 and NGC 4382 (M85). *Astrophys. J.* **599,** 218–236 (2003).
13. Oke, J. B. A new Cassegrain spectrograph for the Hale 5-meter telescope. *Bull. Am. Astron. Soc.* **13,** 509 (1981).
14. Retter, A. & Marom, A. A model of an expanding giant that swallowed planets for the eruption of V838 Monocerotis. *Mon. Not. R. Astron. Soc.* **345,** L25–L28 (2003).
15. Arp, H. C. Novae in the Andromeda nebula. *Astron. J.* **61,** 15–34 (1956).
16. Capaccioli, M., della Valle, M., Rosino, L. & D'Onofrio, M. Properties of the nova population in M31. *Astron. J.* **97,** 1622–1633 (1989).
17. Darnley, M. J. *et al.* Classical novae from the POINT-AGAPE microlensing survey of M31—I. The nova catalogue. *Mon. Not. R. Astron. Soc.* **353,** 571–588 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Frail for discussion and constructive criticism. We also thank the astronomers who maintain the NED database at IPAC and the data archives of the Hubble Space Telescope, the Spitzer Space Telescope and the Chandra X-ray Telescope. Our work has been funded in part by NASA, NSF, the Sylvia and Jim Katz Foundation and the TABASGO Foundation.

Author Contributions S.R.K. recognized the importance of a sub-luminous 'supernova candidate' in M85 (E.O.O. and A.M.S. both separately alerted him to it). The discovery was reported by W.L. and A.V.F. The paper was written mainly by S.R.K. and A.R. E.O.O. and A.R. contributed to the analysis of spectra and the summary figure. Much of the photometry came from the Palomar 60-inch automated telescope. The Palomar 60-inch project is a result of the effort of S.B.C., D.B.F., A.G.-Y., D. S.M. and A.M.S. P.L.C., E.E., J.K. and D.B.S. contributed near-infrared measurements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.R.K. (srk@astro.caltech.edu).

LETTERS

Aggregation and vesiculation of membrane proteins by curvature-mediated interactions

Benedict J. Reynwar[1], Gregoria Illya[1], Vagelis A. Harmandaris[1], Martin M. Müller[1], Kurt Kremer[1] & Markus Deserno[1]

Membrane remodelling[1–5] plays an important role in cellular tasks such as endocytosis, vesiculation and protein sorting, and in the biogenesis of organelles such as the endoplasmic reticulum or the Golgi apparatus. It is well established that the remodelling process is aided by specialized proteins that can sense[4] as well as create[6] membrane curvature, and trigger tubulation[7–9] when added to synthetic liposomes. Because the energy needed for such large-scale changes in membrane geometry significantly exceeds the binding energy between individual proteins and between protein and membrane, cooperative action is essential. It has recently been suggested[10,11] that curvature-mediated attractive interactions could aid cooperation and complement the effects of specific binding events on membrane remodelling. But it is difficult to experimentally isolate curvature-mediated interactions from direct attractions between proteins. Moreover, approximate theories predict repulsion between isotropically curving proteins[12–15]. Here we use coarse-grained membrane simulations to show that curvature-inducing model proteins adsorbed on lipid bilayer membranes can experience attractive interactions that arise purely as a result of membrane curvature. We find that once a minimal local bending is realized, the effect robustly drives protein cluster formation and subsequent transformation into vesicles with radii that correlate with the local curvature imprint. Owing to its universal nature, curvature-mediated attraction can operate even between proteins lacking any specific interactions, such as newly synthesized and still immature membrane proteins in the endoplasmic reticulum.

Far from being a mere outer envelope, lipid bilayer membranes form the basis of many important cellular organelles, such as the endoplasmic reticulum, the Golgi apparatus, or the vesicular transport system. The biological function of these structures often depends on their highly intricate geometry, topology and dynamics, which are actively monitored by the cell. The necessary control is exercised, at least in part, by specialized membrane proteins. These must act cooperatively, as the following simple estimate of the energy requirements shows: at the continuum level, the elastic membrane behaviour is described by a local bending energy per unit area, $E = \frac{1}{2}\kappa(1/R_1 + 1/R_2)^2$, where R_1 and R_2 are the local curvature radii and κ is the bending modulus[16]. For typical phospholipid bilayers, $\kappa \approx 20 k_B T$, where $k_B T \approx 4.1 \times 10^{-21}\,\mathrm{J} \approx 0.6\,\mathrm{kcal\,mol^{-1}}$ is the thermal energy. Creating a spherical vesicle of radius R thus costs about $\frac{1}{2}\kappa(2/R)^2 \times 4\pi R^2 = 8\pi\kappa \approx 500 k_B T$, independent of its radius. This exceeds the typical interaction energy between proteins and also their binding strength to the bilayer by at least an order of magnitude[17].

This energy consideration and the very function of remodelling proteins suggest that specific binding might be complemented by a universal mode of interaction. Such a universal mode can arise because much of the free energy of binding associated with the adsorption of a membrane-curving protein onto a lipid bilayer is stored as elastic bending energy in the membrane. When two membrane-curving proteins approach one another, the bilayer deformations overlap long before any direct interaction occurs. The resulting change in stored bending energy is distance dependent, thus yielding a force. A recent simulation study explicitly posed the question of the cooperative interaction of many such domains[10], and it has been suggested that even without direct interactions membrane-curving proteins might cluster and subsequently tubulate in order to reduce the curvature energy[11]. As two such proteins could share the work needed to bend the membrane and thereby lower the stored elastic energy, one might intuitively expect that the resultant force between them is attractive. Yet existing experimental and theoretical work shows that the sign of the force is anything but obvious.

Experimental and theoretical approaches to quantification of curvature-mediated interactions have proven difficult and inconclusive. Aggregation of proteins[7–9] and colloids[18] has been observed, but the local geometry was not resolved and direct interactions could not be ruled out. Theoretical calculations require the membrane shape of lowest bending energy, but this calls for use of a fourth-order nonlinear partial differential (shape-)equation that can only be solved in very exceptional cases. Approximate linearized solutions for weakly perturbed membranes exist, and suggest that two proteins imposing isotropic curvatures repel[12–15] while attractions require anisotropic curvature imprints[19,20]; however, linearized solutions are not expected to remain valid for strong membrane deformations.

When both experiment and theory encounter difficulties, tailored computer simulations offer an alternative approach, with their unique ability to identify and separate individual contributions to the phenomenon or process of interest. A meaningful simulation of membrane vesiculation calls for model membranes extending in excess of 100 nm and simulation times of the order of milliseconds, which were until recently out of reach for conventional atomistic and many coarse-grained simulations. But these problems are overcome with our recently developed[21] coarse-grained model, which achieves efficient simulation owing to the elimination of explicit solvent molecules (see refs 22 and 23 for current reviews on coarse-grained membrane simulations in general, and ref. 24 for solvent-free models in particular). Like all coarse-grained approaches, the model eliminates atomistic detail and thus cannot be used to explore phenomena dependent on such detail. But on the length scales of tens to hundreds of nanometres relevant to our study, it faithfully reproduces all key properties of self-assembling fluid bilayers, in particular the bending elasticity[21,25]; it thus is a suitable tool for isolating and identifying curvature-mediated interactions between mutually non-interacting local membrane curvers.

Full technical details on the membrane model can be found in ref. 21 and the Supplementary Information. Briefly, it is built from model lipids approximated by three connected beads (see Fig. 1). Bilayer assembly ($\kappa \approx 12\, k_B T$) is triggered by effective tail attractions. The

[1]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany.

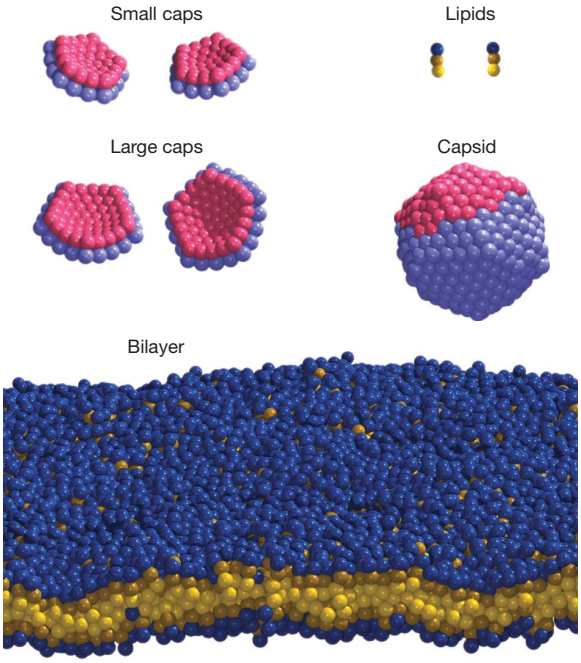


Figure 1 | Illustration of the individual entities used in the simulation. Three-bead lipids with one hydrophilic head-bead and two hydrophobic tail-beads form a flat fluid bilayer spanning the simulation box. Curved caps and full capsids are created from beads of the same size. Only the light blue ones attract the dark blue hydrophilic head-beads of the lipids. None of the beads of caps or capsids attract other caps or capsids.

bead diameter, σ, is set to about 1 nm, to yield an appropriate membrane thickness. The natural simulation timescale is $\tau \approx 15$ ns, based on lipid self-diffusion. The simplified curvature-inducing proteins are curved caps of two different sizes, corresponding to 10% and 16% of a sphere of radius 5.5σ. Their outer surface attracts the hydrophilic lipid beads, thus locally curving the membrane isotropically. The size and degree of deflection of these model particles are comparable to real proteins (see Supplementary Information for details). We also study complete spheres of radius 5σ with 75% of their surface rendered attractive to hydrophilic lipid beads, so we can probe both weak and strong curvature perturbations.

Placing 36 small caps onto a tensionless square membrane with a side length of $\sim 160\sigma \approx 160$ nm (46,080 lipids) results in only weak clustering for the entire $70{,}000\tau$ simulation time (see Supplementary Fig. 1), indicating that any mediated interaction is small compared to the thermal energy. In stark contrast, the large caps behave qualitatively differently (see Fig. 2 and Supplementary Video 1): after initial weak clustering with a protein–protein interaction energy of $\sim 1.3 k_B T$, most caps suddenly aggregate at $\sim 40{,}000\tau$ into a single, almost flat cluster that then rapidly vesiculates within the subsequent $30{,}000\tau \approx 0.5$ ms. Note that throughout this process, individual caps neither touch nor order in a crystalline-like fashion. When using yet larger caps, the aggregation proceeds more rapidly, the aggregates are denser, and the vesicle sizes are smaller because fewer proteins suffice to create them (see Supplementary Video 2). In cellular organelles, this curvature-mediated vesiculation mechanism has to compete against a residual bilayer tension that suppresses vesicle formation beyond a critical size.

The vesiculation pathway observed in our simulations differs fundamentally from scaffolding[3,6] schemes that require direct and specific protein contacts, such as clathrin-dependent endocytosis. Evidently, curvature-mediated interactions alone can induce aggregation and vesiculation (see Supplementary Information for a detailed discussion of the energetics). This effect has not been seen in approximate linearized continuum theories for isotropic membrane-curvers, possibly because the induced deformations are too strong to permit linearization (an approximation to which the sign of an interaction is known to be sensitive[26]). On the other hand, although the attractive interactions seen in our simulations weaken and ultimately vanish as the membrane curvature induced by protein adsorption diminishes, a crossover to the repulsive behaviour predicted by linearized continuum theories could not be identified with statistical significance. However, we note that in this regime other contributions can induce interactions (such as fluctuations[12], or depletion or tilt-mediated[27] forces) and compete with effects arising from the diminishing extent of membrane curvature. We also note that the interaction behaviour for small membrane deformation is only known for large separations, so it is not clear which forces one should expect for small deformations in our system.

Our final simulations use colloidal spheres that could represent viral capsids or nanoparticles. These spheres have a radius comparable to that characterizing the curved caps used before, and 75% of their surface is attractive to the hydrophilic lipid beads. Within the first $2{,}000\tau$ of placing 16 such capsids onto a tensionless square membrane of initial side-length 160σ (see Fig. 3), the membrane contracts to $\sim 140\sigma$ as it coats the attractive part of the colloidal spheres (Fig. 3a to Fig. 3b). After the initial contraction, clustering

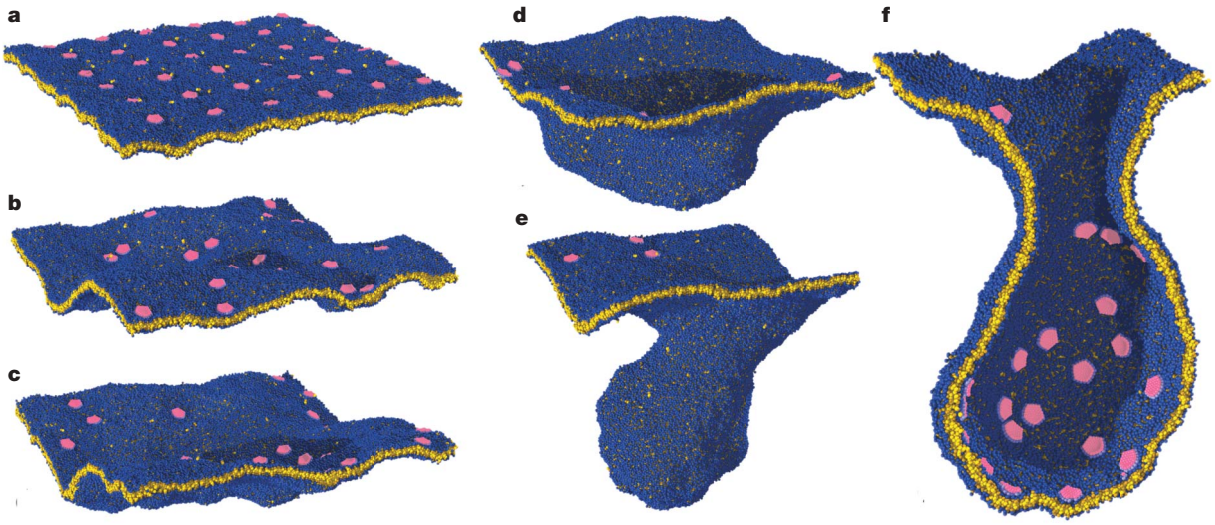


Figure 2 | Successive stages of a vesiculation event driven by 36 large caps on a membrane containing 46,080 lipids. The times of the simulation snapshots are: **a**, 0τ; **b**, $20{,}000\tau$; **c**, $40{,}000\tau$; **d**, $50{,}000\tau$; **e**, $60{,}000\tau$; and **f**, $70{,}000\tau$, the last corresponding to roughly 1 ms.

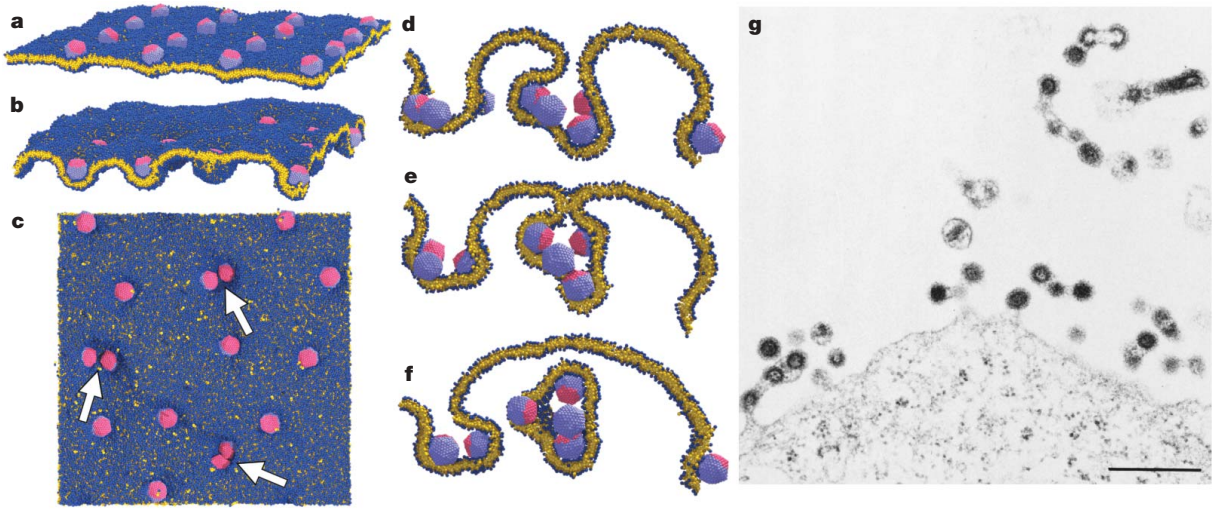


Figure 3 | Attraction and cooperative budding driven by 16 capsids on a membrane containing 40,960 lipids. a–f, A series of simulation snapshots. The times are: **a**, 0τ; **b**, 1,000τ; **c**, 7,000τ; **d**, 16,000τ; **e**, 17,000τ; and **f**, 18,000τ, the last corresponding to roughly 0.3 ms. The arrows in **c** point to formed capsid-pairs. The slices **d–f** indicate cooperative budding, a phenomenon also seen in the electron micrograph (**g**) of late domain mutated MPMV virions (scale bar, 500 nm; reprinted from ref. 28 with permission of authors and publisher; copyright 2003, The American Society for Microbiology).

sets in; it always starts with the formation of pairs (Fig. 3c), and is followed by subsequent tight vesiculation (Fig. 3d–f and Supplementary Video 3). The final multi-capsid structures closely resemble morphologies encountered in the cooperative budding of late domain mutated Mason-Pfizer monkey viruses (MPMV)[28], which lack individual budding activity (Fig. 3g).

To quantify the attraction that induces the spheres to form pairs, we placed two capsids on a membrane and fixed their separation d. As illustrated in Fig. 4, the constraining force needed to maintain this separation revealed a significant attraction, which is strongest around $d = 12\sigma$ and decays very weakly (for very short distances, capsids repel owing to direct contact). System size requirements and slow thermal shape fluctuations make it difficult to determine the interaction behaviour at large distances even for such strong deformers as these capsids, but we nevertheless obtain a total mutual binding energy of $\sim 10 k_B T \approx \kappa$. This value rules out aggregation due to fluctuation effects, and points instead to a true ground state curvature-mediated interaction.

The simulation snapshot in Fig. 4 illustrates that capsids significantly tilt towards each other. We also note the finding of a nonlinear geometry analysis that the net force between the capsids results from a competition between the force associated with the curvature along the direction joining the particles (which drives repulsion) and the force associated with the curvatures perpendicular to it (which drives attraction)[27,29]. Taken together, this information points towards a possible mechanism for capsid attraction: as the colloidal spheres approach each other they flatten the former curvature by tilting, thus enabling the attractive forces associated with the second curvature direction to take over. This effect may be supplemented by a slight 'peeling' of the membrane from the front and back of the capsids, as visible in Fig. 4. Although it is difficult to predict the outcome of this subtle balance analytically, our simulations clearly show that a sufficiently large curvature imprint will result in an overall attraction between membrane-adsorbed proteins. The universal nature of this effect renders it extremely robust, and suggests that cells take advantage of it. In fact, cellular membrane control might even require measures to prevent such omnipresent aggregation.

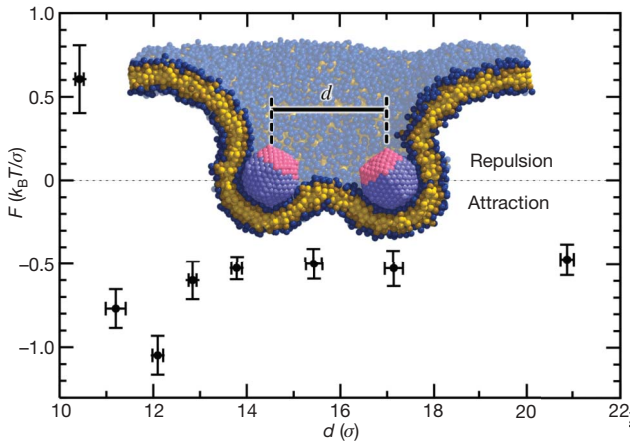


Figure 4 | Force versus distance for two capsids. Negative forces signify attraction. Vertical error bars, ±1 s.e.m.; horizontal error bars, ±1 s.d. Inset, a cross-sectional cut through the membrane profile for a separation of $d = 21\sigma$.

Received 23 October 2006; accepted 11 April 2007.

1. Bannykh, S. I. & Balch, W. E. Membrane dynamics at the endoplasmic reticulum-Golgi interface. *J. Cell Biol.* **138**, 1–4 (1997).
2. Lippincott-Schwartz, J., Roberts, T. H. & Hirschberg, K. Secretory protein trafficking and organelle dynamics in living cells. *Annu. Rev. Cell Dev. Biol.* **16**, 557–589 (2000).
3. McMahon, H. T. & Gallop, J. L. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **438**, 590–596 (2005).
4. Vogel, V. & Sheetz, M. Local force and geometry sensing regulate cell functions. *Nature Rev. Mol. Cell Biol.* **7**, 265–275 (2006).
5. Voeltz, G. K., Prinz, W. A., Shibata, Y., Rist, J. M. & Rapoport, T. A. A class of membrane proteins shaping the tubular endoplasmic reticulum. *Cell* **124**, 573–586 (2006).
6. Zimmerberg, J. & Kozlov, M. M. How proteins produce cellular membrane curvature. *Nature Rev. Mol. Cell Biol.* **7**, 9–19 (2006).
7. Takei, K., Slepnev, V. I., Hauke, V. & De Camilli, P. Functional partnership between amphiphysin and dynamin in clathrin-mediated endocytosis. *Nature Cell Biol.* **1**, 33–39 (1999).
8. Farsad, K. *et al.* Generation of high curvature membranes mediated by direct endophilin bilayer interactions. *J. Cell Biol.* **155**, 193–200 (2001).
9. Itoh, T. *et al.* Dynamin and the actin cytoskeleton cooperatively regulate plasma membrane invagination by BAR and F-BAR proteins. *Dev. Cell* **9**, 791–804 (2005).
10. Blood, P. D. & Voth, G. A. Direct observation of Bin/amphiphysin/Rvs (BAR) domain-induced membrane curvature by means of molecular dynamics simulations. *Proc. Natl Acad. Sci. USA* **103**, 15068–15072 (2006).
11. Antonny, B. Membrane deformation by protein coats. *Curr. Opin. Cell Biol.* **18**, 386–394 (2006).
12. Goulian, M., Bruinsma, R. & Pincus, P. Long-range forces in heterogeneous fluid membranes. *Europhys. Lett.* **22**, 145–150 (1993).

13. Weikl, T. R., Kozlov, M. M. & Helfrich, W. Interaction of conical membrane inclusions: Effect of lateral tension. *Phys. Rev. E* **57,** 6988–6995 (1998).
14. Bartolo, D. & Fournier, J.-B. Elastic interaction between ''hard'' or ''soft'' pointwise inclusions on biological membranes. *Eur. Phys. J. E* **11,** 141–146 (2003).
15. Kim, K. S., Neu, J. & Oster, G. Curvature-mediated interactions between membrane proteins. *Biophys. J.* **75,** 2274–2291 (1998).
16. Helfrich, W. Elastic properties of lipid bilayers — Theory and possible experiments. *Z. Naturforsch. C* **28,** 693–703 (1973).
17. Zimmerberg, J. & McLaughlin, S. Membrane curvature: How BAR domains bend bilayers. *Curr. Biol.* **14,** R250–R252 (2004).
18. Koltover, I., Rädler, J. O. & Safinya, C. R. Membrane mediated attraction and ordered aggregation of colloidal particles bound to giant phospholipid vesicles. *Phys. Rev. Lett.* **82,** 1991–1994 (1999).
19. Chou, T., Kim, K. S. & Oster, G. Statistical thermodynamics of membrane bending-mediated protein-protein attractions. *Biophys. J.* **80,** 1075–1087 (2001).
20. Fournier, J.-B., Dommersnes, P. G. & Galatola, P. Dynamin recruitment by clathrin coats: A physical step? *C. R. Biol.* **326,** 467–476 (2003).
21. Cooke, I. R., Kremer, K. & Deserno, M. Tunable generic model for fluid bilayer membranes. *Phys. Rev. E* **72,** 011506 (2005).
22. Müller, M., Katsov, K. & Schick, M. Biological and synthetic membranes: What can be learned from a coarse-grained description? *Phys. Rep.* **434,** 113–176 (2006).
23. Venturoli, M., Sperotto, M. M., Kranenburg, M. & Smit, B. Mesoscopic models of biological membranes. *Phys. Rep.* **437,** 1–54 (2006).
24. Brannigan, G., Lin, L. C.-L. & Brown, F. L. H. Implicit solvent simulations for biomembranes. *Eur. Biophys. J.* **35,** 104–124 (2006).
25. Harmandaris, V. A. & Deserno, M. A novel method for measuring the bending rigidity of model lipid membranes by simulating tethers. *J. Chem. Phys.* **125,** 204905 (2006).
26. Goulding, D. & Hansen, J.-P. Attraction between like-charged colloidal particles induced by a surface: A density functional analysis. *Europhys. Lett.* **46,** 407–413 (1999).
27. Müller, M. M., Deserno, M. & Guven, J. Interface mediated interactions between particles — a geometrical approach. *Phys. Rev. E* **72,** 061407 (2005).
28. Gottwein, E. *et al.* The Mason-Pfizer monkey virus PPPY and PSAP motifs both contribute to virus release. *J. Virol.* **77,** 9474–9485 (2003).
29. Müller, M. M., Deserno, M. & Guven, J. Geometry of surface mediated interactions. *Europhys. Lett.* **69,** 482–488 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We enjoyed discussions with I. R. Cooke, Ch. Peter, E.-K. Sinner, J. Guven, H.-G. Kräusslich and all members of the ESPResSo team at the MPI-P. B.J.R. acknowledges financial support from the collaborative research centre 'From single molecules to nanoscopically structured materials' and M.D. from an Emmy Noether fellowship, both of the Deutsche Forschungsgemeinschaft. A grant for computer time within the DEISA programme is also acknowledged.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.D. (deserno@mpip-mainz.mpg.de) or K.K. (kremer@mpip-mainz.mpg.de).

LETTERS

Intense hurricane activity over the past 5,000 years controlled by El Niño and the West African monsoon

Jeffrey P. Donnelly[1] & Jonathan D. Woodruff[1]

The processes that control the formation, intensity and track of hurricanes are poorly understood[1]. It has been proposed that an increase in sea surface temperatures caused by anthropogenic climate change has led to an increase in the frequency of intense tropical cyclones[2,3], but this proposal has been challenged on the basis that the instrumental record is too short and unreliable to reveal trends in intense tropical cyclone activity[4]. Storm-induced deposits preserved in the sediments of coastal lagoons offer the opportunity to study the links between climatic conditions and hurricane activity on longer timescales, because they provide centennial- to millennial-scale records of past hurricane landfalls[5–8]. Here we present a record of intense hurricane activity in the western North Atlantic Ocean over the past 5,000 years based on sediment cores from a Caribbean lagoon that contain coarse-grained deposits associated with intense hurricane landfalls. The record indicates that the frequency of intense hurricane landfalls has varied on centennial to millennial scales over this interval. Comparison of the sediment record with palaeo-climate records indicates that this variability was probably modulated by atmospheric dynamics associated with variations in the El Niño/Southern Oscillation and the strength of the West African monsoon, and suggests that sea surface temperatures as high as at present are not necessary to support intervals of frequent intense hurricanes. To accurately predict changes in intense hurricane activity, it is therefore important to understand how the El Niño/Southern Oscillation and the West African monsoon will respond to future climate change.

At present there is significant debate about the cause of observed multi-decadal variability of hurricanes in the North Atlantic (for example, see refs 2, 4). To detect long-term patterns in tropical cyclone activity, reliable proxy reconstructions that extend back before the instrumental record are needed. To examine the centennial- and millennial-scale variability of Caribbean hurricane activity and to assess potential climate forcing we reconstruct the history of hurricane-induced overwash events from Laguna Playa Grande (LPG), Vieques, Puerto Rico.

The island of Vieques is located in the northeastern Caribbean Sea (Fig. 1) and is extremely vulnerable to hurricanes. LPG is a hypersaline, backbarrier lagoon separated from the Caribbean Sea by a wave-dominated, sandy barrier 80 m wide and 2–3 m high. The barrier is stabilized on either end by rocky headlands[9] and anchored below by beach rock[10]. Tidal variability is modest (mean range 0.24 m), which minimizes the influence of tidal currents and inlet dynamics. In addition, the relatively slow rates of sea-level rise over the past 6,000 years in the region[11] and the steep topography and bathymetry contribute to barrier stability.

Cores collected from the site contain several metres of organic-rich silt interbedded with coarse-grained event layers comprised of a mixture of siliciclastic sand and calcium carbonate shells and shell fragments. These layers are the result of marine flooding events overtopping or breaching the barrier and transporting these barrier and nearshore sediments into the lagoon. Patterns of coarse-grained event deposits are consistent among all cores (Fig. 2, Supplementary Fig. 1). To determine which historical events left coarse-grained layers at LPG, we developed a detailed age model for the upper 20 cm of LPG12 (Fig. 2). Three coarse-grained deposits are evident in the sediments deposited within the past 100 years. These layers are consistent with three of the most intense hurricanes to strike Vieques over this interval. Seven hurricanes passed within 50 km of the site between 1900 and 2006. Of these, the dates for the two most extreme storms (hurricanes San Felipe in 1928 (category 5) and Hugo in 1989 (category 4)) are consistent with the age of two of the three layers

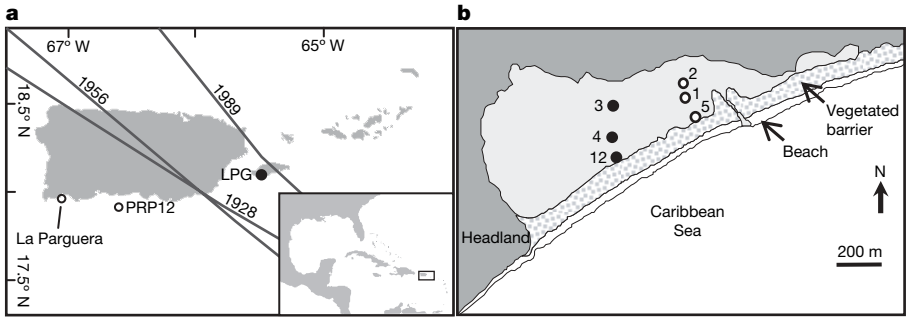


Figure 1 | Site map and core locations. a, Map of Puerto Rico with inset map of the tropical Western Atlantic. The location of LPG on the southeastern coast of Vieques is noted with a solid circle. Tracks of the hurricanes mentioned in the text are noted. The location of Puerto Rico (box) is indicated in the inset. Locations of SST reconstructions from La Parguera[17] and PRP 12[16] are noted. **b**, Map of LPG showing core locations (circles). The locations of the cores (LPG12, LPG4 and LPG3) presented in Fig. 2 are noted with solid circles. Cores LPG5, LPG1 and LPG2 included in Supplementary Fig. 1 are also noted.

[1]Coastal Systems Group, Woods Hole Oceanographic Institution, 360 Woods Hole Road, Woods Hole, Massachusetts 02543, USA.

observed in LPG12 (Fig. 2). Of the less intense storms (categories 1 and 2) only hurricane Betsy (1956) correlates well with the third deposit. However, an analysis of wind damage in eastern Puerto Rico for this particular storm indicates wind speeds more consistent with category 3 intensity[12]. In contrast to the three layers in LPG12, only two layers are preserved in the upper 15 cm of LPG4. The same pattern is evident in cores collected along the easternmost transect (Fig. 1 and Supplementary Fig. 1). The more distal coarse-grained layers in LPG4 were probably deposited by the two most intense hurricanes in the past 100 years (hurricanes San Felipe in 1928 and Hugo in 1989).

Areas of the lagoon adjacent to the barrier (for example, LPG5 (Fig. 2) and LPG12 (Supplementary Fig. 1)) are more likely to experience localized breaching associated with less intense events and are also more susceptible to erosion and truncation of the sediment record during overwash. Conversely, coarse-grained sediments do not always reach the most distal locations (for example, LPG3 (Fig. 2) and LPG2 (Supplementary Fig. 1)) during extreme events, and as a result these areas provide an incomplete record (Fig. 2). The coarse-grained deposits in central locations of the lagoon (for example, LPG4 (Fig. 2) and LPG1 (Supplementary Fig. 1)) provide a relatively complete record of the most intense hurricane (category 4 and greater) strikes, because only these extreme events are capable of producing storm surges high enough to overtop the entire length of the barrier and carry and deposit coarse-grained layers to these locations. Storm-induced deposits within LPG4 reveal large fluctuations in the frequency of intense hurricanes (Fig. 3a). On the basis of our age model (Supplementary Fig. 2) an interval of relatively frequent intense hurricane strikes at Vieques is evident between 5,400 and 3,600 calendar years before present (yr BP, where present is defined as 1950 AD by convention), with the exception of a short-lived quiescent interval between approximately 4,900 and 5,050 yr BP. Following this relatively active period is an interval of relatively few extreme coastal flooding events persisting from 3,600 until roughly 2,500 yr BP. Evidence of another relatively active interval of intense hurricane strikes is evident between 2,500 and approximately 1,000 yr BP. The interval from 1,000 to 250 yr BP was relatively quiescent with evidence of only one prominent event occurring around 500 yr BP. A relatively active regime has resumed since about 250 yr BP (1700 AD).

Evidence of hurricane landfalls in New York indicates periods of activity similar to those of Vieques over the past 2,500 years[13]. In addition, sediment-derived records of intense hurricanes from the Gulf coast also indicate a relatively quiescent interval beginning about 1,000 years ago[5]. The synchronous transition from frequent to infrequent hurricane landfalls in these three regions indicates that a North-Atlantic-wide decrease in hurricane activity occurred about 1,000 years ago and was not simply a change in prevailing hurricane tracks away from the Gulf coast, as has previously been suggested[5,14].

Warm sea surface temperatures (SSTs) in the tropical North Atlantic are thought to be a key ingredient for fuelling intense hurricanes[1,15] and are at the centre of the debate over the impact of global warming on tropical cyclone activity. Unfortunately, few reconstructions of SST spanning the past 5,000 years from the main development region (MDR) for hurricane formation (Supplementary Fig. 3) are available. However, SST reconstructions from off Puerto Rico (PRP12[16] and La Parguera[17]; Fig. 1a) are probably good proxies for the MDR (Supplementary Fig. 3). The PRP12 reconstruction indicates that summer SSTs in the tropical North Atlantic have generally been cooler than at present, varying by as much as 2 °C (roughly 26–28 °C) over the past 2,000 years (Fig. 3b)[16]. In addition, coral-based SST reconstructions from La Parguera, Puerto Rico (Fig. 1a), indicate that mean annual Little Ice Age (250–135 yr BP or 1700–1815 AD) SSTs were 2–3 °C cooler than they are now (Fig. 3b)[17]. Despite cooler Little Ice Age SSTs in the region, the sediment record from LPG and New York[13] indicates an increase in intense hurricane landfalls since about 1700 AD (250 yr BP) (Fig. 3b).

Historical records from Puerto Rico also suggest an increase in severe hurricane damage in the 18th and 19th centuries. Only three storms are documented as resulting in severe damage (≥F2 on the Fujita scale) in Puerto Rico between 1550 and 1700 AD, while at least sixteen severe hurricanes affected Puerto Rico between 1700 and 1850 AD[12]. Although the historic archives may be less complete during the early part of these records and so some hurricanes may have gone unrecorded, sediment-based reconstructions are unaffected by this type of biasing. Therefore, the good agreement between the sediment-based reconstructions and the historic archives strongly suggests that the frequency of intense hurricanes increased at around 1700 AD. In addition, an analysis of Caribbean hurricanes documented in Spanish archives indicates that 1766–1780 was one of the most active intervals in the period between 1500 and 1800 AD (ref. 18), when tree-ring-based reconstructions indicate a negative (cooler) phase of the Atlantic Multidecadal Oscillation[19]. Furthermore, the sediment record from LPG indicates that an interval of relatively frequent intense hurricane strikes persisted for over a millennium (2,500 to 1,000 years ago) despite cooler-than-modern SSTs. Thus the information available suggests that tropical Atlantic SSTs were probably not the principal

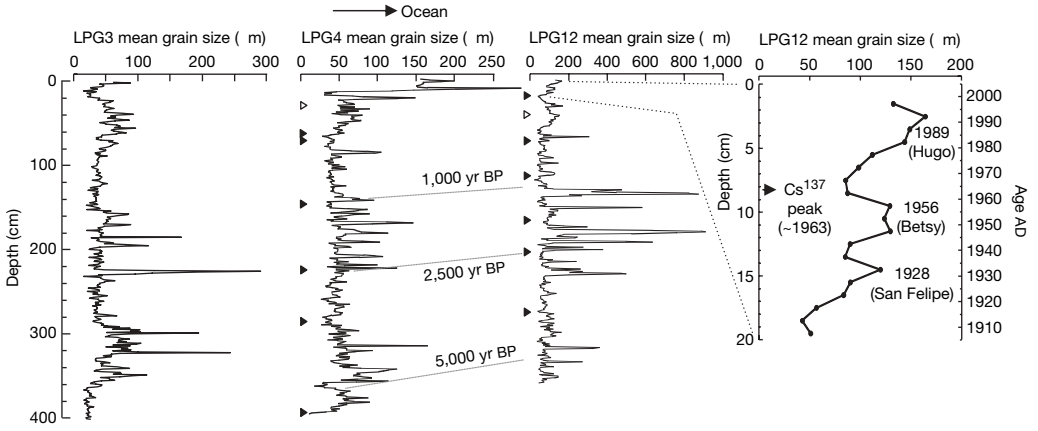


Figure 2 | Bulk grain-size data revealing storm-induced deposits. The mean grain-size scale is set at 250 and 300 μm for LPG4 and LPG3, respectively, because the coarse-grained layers are generally finer in these more distal locations. Solid arrows represent the depth of radiocarbon-dated samples from LPG4 and LPG12. The depth of abrupt increase in Ti and Fe associated with land clearance at approximately 1840 AD (110 yr BP) is noted with an open triangle. Dashed grey lines indicate depths of equal age between LPG4 and LPG12, based on the age models presented in Supplementary Fig. 2. The enlarged upper 20 cm mean grain-size plot of LPG12 (far right) with an age model based on an accumulation rate of 2 mm yr^{-1} is derived from Cs^{137} data and evidence of land clearance (~1840 AD). Deposits attributable to documented hurricanes are noted. The arrow above plots indicates the direction of the ocean relative to core sites.

driver of intense hurricane activity over the past several millennia; however, more high-resolution records of SSTs, including depth of the mixed layer, are required to address the role of SSTs on intense hurricane activity over this period adequately.

Studies relying on recent climatology indicate that North Atlantic hurricane activity is greater during La Niña years and suppressed during El Niño years[20,21], due primarily to increased vertical wind shear in strong El Niño years hindering hurricane development. A comparison between LPG4 and a proxy record of El Niño events from Laguna Pallcacocha, Ecuador[22], suggests that the evolution of El Niño/Southern Oscillation (ENSO) variability has also played a key part in governing Atlantic intense tropical cyclone activity for much of the past 5,000 years (Fig. 3). For example, intervals of frequent intense hurricane strikes at LPG (for example, ~2,500 to 1,000 yr BP, 3,600 to 4,400 yr BP, and 250 yr BP to present) correspond primarily to periods with relatively few El Niño events. Conversely, periods with more frequent, strong El Niño events generally correspond to periods with fewer intense hurricane strikes on Vieques (for example, ~3,600 to 2,500 yr BP and 1,000 to 250 yr BP) (Fig. 3c). A possible exception to this correlation is the interval between 4,600 and 5,000 yr BP; however, a small (~100–200 years) shift within the uncertainty range of the age model in either record would also result in the lull in El Niño events corresponding to the active hurricane interval here.

In addition to the El Niño record there is also a strong correspondence between a precipitation record from Lake Ossa, West Cameroon (Fig. 3d)[23], and the record of intense hurricane activity from Vieques. Intervals of increased precipitation (thought to result from more frequent convective storms) in tropical Africa correspond to times of increased frequency of intense hurricanes recorded at LPG. Conversely, less convective storminess in tropical Africa appears to be associated with relatively few intense hurricanes in the Western Atlantic. The amount of precipitation in tropical Africa is probably related to the strength of the West African monsoon. This correlation between tropical African precipitation and North Atlantic hurricanes is consistent with recent findings[24] linking periods of increased hurricane activity in the middle of the 20th century with increased monsoonal strength in Africa and a well-developed African easterly jet. Increased cyclonic vorticity in the MDR results from a well-developed African easterly jet. During intervals of increased monsoonal strength (with a well-developed African easterly jet) and cool ENSO phase, African easterly disturbances (waves) pass through a region of enhanced cyclonic vorticity, warm SSTs, and low vertical shear, enhancing the development of hurricanes in the central and western portions of the MDR.

Increases in precipitation in tropical Africa are a likely positive feedback mechanism contributing to the formation and enhancement of the African easterly jet by increasing the soil moisture gradient[25]. The negative correlation between precipitation proxies in West Cameroon and Ecuador may also point to ENSO modulation of the West African monsoon. In fact, El Niño events combined with

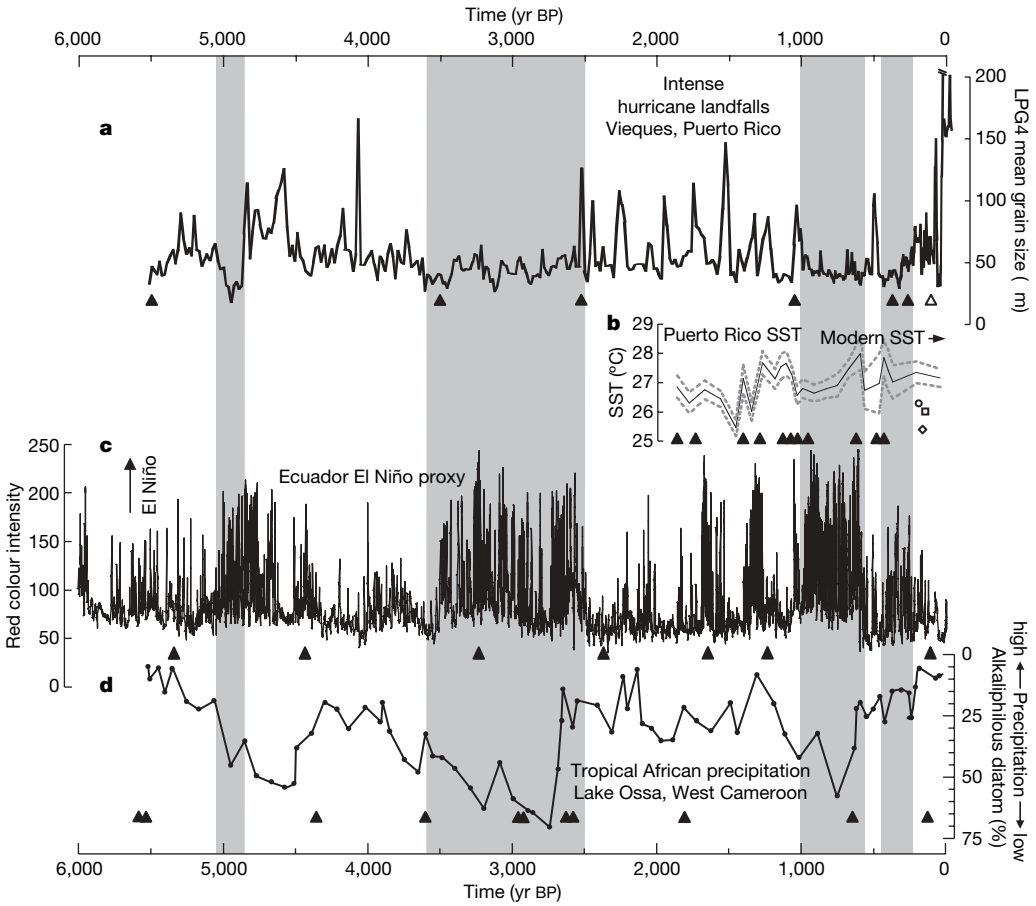


Figure 3 | Comparison of the intense hurricane record from LPG with other climate records. **a**, Mean bulk grain-size record from LPG4. Intervals of relatively few intense-hurricane-induced layers in all cores are noted with shading. **b**, The thin line with the 2σ uncertainty envelope (dashed lines) is a reconstruction of summer SSTs off Puerto Rico[16] (core PRP12) and coral-based reconstruction of mean annual SSTs from La Parguera, Puerto Rico[17], are noted: 26.2 °C for 1700–1705 AD (circle), 25.3 °C for 1780–1785 AD (diamond), and 26.0 °C for 1800–1805 AD (square). The modern mean annual SST is noted with an arrow. **c**, El Niño proxy reconstruction from Laguna Pallcacocha, Ecuador[22]. Peaks in red colour intensity are documented as allochthonous material washed into the lake primarily during strong El Niño events. **d**, Changes in precipitation in West Cameroon inferred from alkaliphilous diatoms (thriving in alkaline conditions) from Lake Ossa[23]. Radiocarbon age control points are noted with black arrows below all panels.

negative SST anomaly in the eastern equatorial Atlantic have been linked to drought in western Africa[26]. However, controls on eastern equatorial Atlantic SST fluctuations independent of ENSO may also have played an important part in modulating the intensity of the West African monsoon over the Holocene epoch[27].

A coherent pattern of climate change over the past 5,000 years appears to have modulated intense hurricane activity in the north-eastern Caribbean. The evolution of ENSO variability over the past several millennia probably played an important part in controlling the frequency of intense hurricanes in the Caribbean and perhaps the entire North Atlantic Basin, with intervals of fewer strong El Niño events resulting in less vertical wind shear over the tropical North Atlantic and more favourable conditions for intense hurricane development. In addition, variations in the West African monsoon and African easterly jet probably also play a critical role in modulating the frequency of North Atlantic intense hurricanes, with increases in convective storms over tropical Africa leading to stronger easterly waves moving into the tropical North Atlantic. Given the increase of intense hurricane landfalls during the later half of the Little Ice Age, tropical SSTs as warm as at present are apparently not a requisite condition for increased intense hurricane activity. In addition, the Caribbean experienced a relatively active interval of intense hurricanes for more than a millennium when local SSTs were on average cooler than modern. These results suggest that in addition to fluctuations in tropical Atlantic SST, changes in atmospheric dynamics tied to ENSO and the West African monsoon also act to modulate intense hurricane activity on centennial and millennial timescales. A better understanding of how these climate patterns will vary in the future is therefore required if we are to predict changes in intense hurricane activity accurately.

METHODS

Cores were collected using a Vohnout/Colinvaux piston corer in 5-cm-diameter polycarbonate barrels. Short 10-cm-diameter push cores were taken at select core locations to capture the sediment/water interface better and to provide adequate material for radio-isotopic analyses. These push cores were extruded in the field and sampled every 0.5 cm. Measurements of the activity of Cs^{137} (a product of atmospheric nuclear weapons testing) were conducted using a high-resolution gamma detector. The locations for all coring sites were determined using a handheld GPS unit which provided a horizontal accuracy of 3 to 6 m. Sediment cores were split in the laboratory and selected core halves were run through a non-destructive Itrax core scanner to obtain millimetre- to submillimetre-resolution X-ray fluorescence measurements of the sediment's elemental composition, on the basis of methods described in ref. 28. Bulk grain-size analysis was conducted on contiguous 1-cm samples using a Beckman-Coulter LS13320 laser diffraction particle-size analyser. As the bulk mean grain-size data represent siliciclastic, organic and calcareous material of varying densities, the relative magnitude of events cannot be directly inferred by comparing values for individual coarse layers. Samples of wood, seeds and shells were radiocarbon-dated at the National Ocean Sciences AMS Facility at Woods Hole Oceanographic Institution. The resulting radiocarbon ages were calibrated to calendar years using the IntCal04[29] and Marine04[30] calibration data sets (Supplementary Table 1).

Received 21 November 2006; accepted 10 April 2007.

1. Goldenberg, S. B., Landsea, C. W., Mestas-Nunez, A. M. & Gray, W. M. The recent increase in Atlantic hurricane activity: Causes and implications. *Science* **293,** 474–479 (2001).
2. Emanuel, K. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **436,** 686–688 (2005).
3. Webster, P. J., Holland, G. J., Curry, J. A. & Chang, H.-R. Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science* **309,** 1844–1846 (2005).
4. Landsea, C. W., Harper, B. A., Hoarau, K. & Knaff, J. A. Can we detect trends in extreme tropical cyclones. *Science* **313,** 452–454 (2006).
5. Liu, K. B. & Fearn, M. L. Reconstruction of prehistoric landfall frequencies of catastrophic hurricanes in northwestern Florida from lake sediment records. *Quat. Res.* **54,** 238–245 (2000).
6. Donnelly, J. P. *et al.* 700 yr sedimentary record of intense hurricane landfalls in southern New England. *Geol. Soc. Am. Bull.* **113,** 714–727 (2001).
7. Donnelly, J. P., Butler, J., Roll, S., Wengren, M. & Webb, T. A backbarrier overwash record of intense storms from Brigantine, New Jersey. *Mar. Geol.* **210,** 107–121 (2004).
8. Donnelly, J. P. Evidence of past intense tropical cyclones from backbarrier salt pond sediments: A case study from Isla de Culebrita, Puerto Rico, USA. *J. Coast. Res.* **SI42,** 201–210 (2005).
9. Roy, P. S., Cowell, P. J., Ferland, M. A. & Thom, B. G. in *Coastal Evolution: Late Quaternary Shoreline Morphodynamics* (eds Carter, R. W. G. & Woodroffe, C. D.) 121–186 (Cambridge Univ. Press, Cambridge, UK, 1994).
10. Cooper, J. A. G. Beachrock formation in low latitudes—implications for coastal evolutionary models. *Mar. Geol.* **98,** 145–154 (1991).
11. Lighty, R. G., Macintyre, I. G. & Struckenrath, R. Acropora Palmata reef framework: a reliable indicator of sea level in the western Atlantic for the past 10,000 years. *Coral Reefs* **1,** 125–130 (1982).
12. Boose, E. R., Serrano, M. I. & Foster, D. R. Landscape and regional impacts of hurricanes in Puerto Rico. *Ecol. Monogr.* **74,** 335–352 (2004).
13. Scileppi, E. & Donnelly, J. P. Sedimentary evidence of hurricane strikes in western Long Island, New York. *Geochem. Geophys. Geosyst.* (in the press).
14. Elsner, J. B., Liu, K. B. & Kocher, B. Spatial variations in major US hurricane activity: Statistics and a physical mechanism. *J. Clim.* **13,** 2293–2305 (2000).
15. Emanuel, K. The dependence of hurricane intensity on climate. *Nature* **326,** 483–485 (1987).
16. Nyberg, J., Malmgren, B. A., Kuijpers, A. & Winter, A. A centennial-scale variability of tropical North Atlantic surface hydrography during the late Holocene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **183,** 25–41 (2002).
17. Winter, A., Ishioroshi, H., Watanabe, T., Oba, T. & Christy, J. Caribbean sea surface temperatures: two to three degrees cooler than present during the Little Ice Age. *Geophys. Res. Lett.* **27,** 3365–3368 (2000).
18. Garcia-Herrera, R., Gimeno, L., Ribera, P. & Hernandez, E. New records of Atlantic hurricanes from Spanish documentary sources. *J. Geophys. Res.* **110,** 1–7 (2005).
19. Gray, S. T., Graumlich, L. J., Betancourt, J. L. & Pederson, G. T. A tree-ring-based reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D. *Geophys. Res. Lett.* **31,** 1–4 (2004).
20. Gray, W. M. Atlantic seasonal hurricane frequency. Part I: El Nino and 30 mb quasi-biennial oscillation influences. *Mon. Weath. Rev.* **112,** 1649–1668 (1984).
21. Bove, M. C., Elsner, J. B., Landsea, C. W., Niu, X. F. & O'Brien, J. J. Effect of El Nino on US landfalling hurricanes, revisited. *Bull. Am. Meteorol. Soc.* **79,** 2477–2482 (1998).
22. Moy, C. M., Seltzer, G. O., Rodbell, D. T. & Anderson, D. M. Variability of El Nino/Southern Oscillation activity at millennial timescales during the Holocene epoch. *Nature* **420,** 162–165 (2002).
23. Nguetsop, V. F., Servant-Vildary, S. & Servant, M. Late Holocene climatic changes in west Africa, a high resolution diatom record from equatorial Cameroon. *Quat. Sci. Rev.* **23,** 591–609 (2004).
24. Bell, G. D. & Chelliah, M. Leading tropical modes associated with interannual and multidecadal fluctuations in North Atlantic hurricane activity. *J. Clim.* **19,** 590–612 (2006).
25. Cook, K. H. Generation of the African easterly jet and its role in determining West African precipitation. *J. Clim.* **12,** 1165–1184 (1999).
26. Janicot, S., Harzallah, A., Fontaine, B. & Moron, V. West African monsoon dynamics and eastern equatorial Atlantic and Pacific SST anomalies (1970–88). *J. Clim.* **11,** 1874–1882 (1998).
27. Weldeab, S., Schneider, R. R., Kolling, M. & Wefer, G. Holocene African droughts relate to eastern equatorial Atlantic cooling. *Geology* **33,** 981–984 (2005).
28. Croudace, I. W., Rindby, A. & Rothwell, R. G. ITRAX: description and evaluation of a new X-ray core scanner. In *New Techniques in Sediment Core Analysis* (ed. Rothwell, R. G.) *Geol. Soc. Lond. Spec. Publ.* **267,** 51–63 (2006).
29. Reimer, P. J. *et al.* IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* **46,** 1029–1058 (2004).
30. Hughen, K. A. *et al.* Marine04 marine radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* **46,** 1059–1086 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding for this research was provided by the National Science Foundation, the Risk Prediction Initiative, the National Geographic Society, the Coastal Ocean Institute at Woods Hole Oceanographic Institution, and the Andrew W. Mellon Foundation Endowed Fund for Innovative Research. We are grateful to E. Bryant, E. Scileppi, J. Tierney, and A. Jovanovic who assisted with the field and laboratory work. E. Uchupi and P. Lane provided advice and D. Oppo, J. Russell, T. Webb III, K. Emanuel and L. Giosan made suggestions for improving this manuscript. This is a contribution of IGCP 495—'Holocene land-ocean interactions: driving mechanisms and coastal responses'.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.P.D. (jdonnelly@whoi.edu).

LETTERS

Evolution of cooperation in a finite homogeneous graph

Peter D. Taylor[1], Troy Day[1] & Geoff Wild[1]

Recent theoretical studies of selection in finite structured populations[1–7] have worked with one of two measures of selective advantage of an allele: fixation probability and inclusive fitness. Each approach has its own analytical strengths, but given certain assumptions they provide equivalent results[1]. In most instances the structure of the population can be specified by a network of nodes connected by edges (that is, a graph)[8–10], and much of the work here has focused on a continuous-time model of evolution, first described by ref. 11. Working in this context, we provide an inclusive fitness analysis to derive a surprisingly simple analytical condition for the selective advantage of a cooperative allele in any graph for which the structure satisfies a general symmetry condition ('bi-transitivity'). Our results hold for a broad class of population structures, including most of those analysed previously, as well as some for which a direct calculation of fixation probability has appeared intractable. Notably, under some forms of population regulation, the ability of a cooperative allele to invade is seen to be independent of the nature of population structure (and in particular of how game partnerships are specified) and is identical to that for an unstructured population. For other types of population regulation our results reveal that cooperation can invade if players choose partners along relatively 'high-weight' edges.

In evolutionary biology an individual's behaviour is termed 'social' if it affects the reproductive success of other individuals in the population, and one of the most enduring puzzles is the existence of social behaviours such as cooperation and altruism. Hamilton's[12] pioneering work on inclusive fitness demonstrated that, if individuals tend to interact with genetic relatives, then such behaviours can often evolve. Interactions with relatives might occur as a result of kin recognition, but they might also occur simply as a result of limited dispersal[13–16].

Recent studies in evolutionary game theory[8–10] have provided some interesting new results on the evolution of cooperation, by describing patterns of interaction among individuals in terms of graphs. A graph is simply a set of nodes (representing the individuals), each of which is connected to other nodes with edges (Box 1). These edges provide the relationships among individuals and are of two types: the dispersal patterns of offspring, and the patterns of social interaction among individuals. The probability of fixation of a cooperative allele can then be determined, as a function of the structure of the graph, so that we might understand how different kinds of networks of interactions affect the evolution of social behaviour[8–10]. Here we demonstrate that Hamilton's notion of inclusive fitness[12] provides a natural way to understand evolution on such graphs, and that it provides simple analytical conditions for the evolution of any trait (including cooperation) for a large class of graphs. The primary process at work in such systems can thus be viewed as a case of interactions among related individuals as a result of limited dispersal. A schematic summarizing our main results is provided as Supplementary Fig. 1.

In the graphs considered here, the edges represent patterns of dispersal and social interactions given by the weights d_{ij} and e_{ij} respectively. Specifically, the weights d_{ij} determine (in a way made precise below) the probability that a birth at node i replaces the individual at node j. Furthermore, individuals i and j interact at each time step with probability e_{ij}. We assume $d_{ii} = e_{ii} = 0$, $\sum_j d_{ij} = \sum_j e_{ij} = 1$, and $d_{ij} = d_{ji}$ and $e_{ij} = e_{ji}$. We suppose that the fecundity $f_i = F(X_i, Y_i)$ of individual i (where F is the fecundity function specified by the model) is determined by its own phenotype X_i and by the average phenotype $Y_i = \sum_j e_{ij}X_j$ of the individuals with whom it interacts.

We assume that individuals are haploid with one of two alleles A or B, and offspring are identical to parents except for a small symmetric probability of mutation. We consider two versions of fecundity selection. In the BD process, births are allocated to the population at a fixed, fitness-independent rate and are given to individual i with relative probability f_i, replacing a neighbour j of i with probability d_{ij}. In the DB process, individuals die at a fixed, fitness-independent rate and a death at node j is replaced by an offspring from node i with relative probability $f_i d_{ij}$. In both cases, the population size remains constant. Take note that f_i measures fecundity, but this is only one component of fitness, the other being mortality.

Our results apply to an arbitrary fecundity function $F(X, Y)$ provided that the 'effect' of the A allele is both small and additive. For example, our results apply to the general matrix game[10] $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ (where the first row gives the pay-off to an A player against an A or B opponent, and the second row gives the same pay-offs to a B player), provided both that the a_{hk} values are small and that $a_{11} + a_{22} = a_{12} + a_{21}$. We work here with the 'cost–benefit' matrix $\begin{bmatrix} b-c & -c \\ b & 0 \end{bmatrix}$. A second example is found in Box 2.

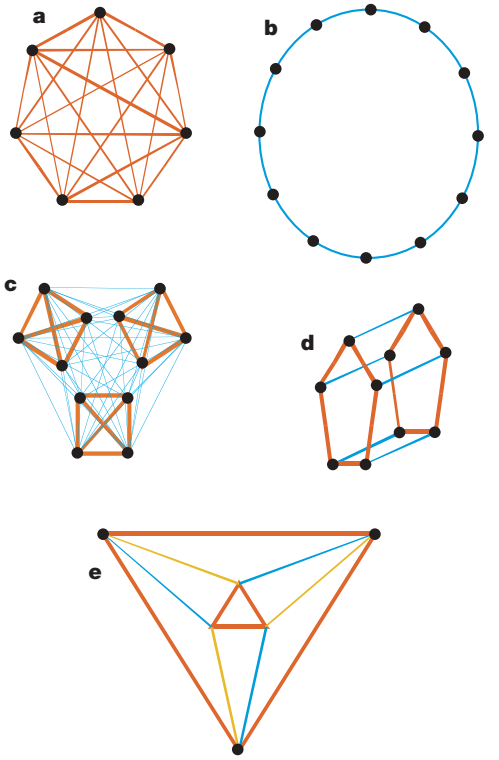
We consider graphs with a large amount of internal symmetry as described by an isomorphism. An isomorphism T of a graph is a bijection of the node set that preserves the dispersal and interaction parameters, that is, $d_{T(i)T(j)} = d_{ij}$ and $e_{T(i)T(j)} = e_{ij}$. A graph is called (node) transitive if, for every ordered pair of nodes (i, j), there is an isomorphism T for which $T(i) = j$; it is called (node) bi-transitive if, for every ordered pair of nodes (i, j), there is an isomorphism T for which $T(i) = j$ and $T(j) = i$. The graphs given in Box 1 are all transitive and all but panel e are bi-transitive.

Our objective is to measure the selective advantage of the allele A. We begin in the neutral population ($c = b = 0$) in which A and B are equally fit, and ask: what is the effect of increasing b and c above 0? We focus attention on three different selective measures. First, the change in b and c will cause the expected long-term frequency of A (E(A); under mutation-selection balance) to differ from E(B), the long-term frequency of B, and we might use that difference as a measure. Second, we define the fixation probability ρ_A of A (ρ_B of B) to be

[1]Department of Mathematics and Statistics, Queen's University Kingston, Ontario K7L 3N6, Canada.

Box 1 | Transitive graphs

Roughly speaking, a graph is transitive if it globally 'looks the same' from any node; that is, if you placed an individual at a node and then blindfolded him and perhaps moved him to another node, he would be unable to tell, using only information about total configuration of edges and their weights, whether or not he had been moved. A graph is bi-transitive if it 'looks the same' from any pair of nodes. That is, if you placed two individuals at any two nodes and then blindfolded them and perhaps interchanged them, they would be unable to tell whether or not they had been interchanged. In the examples shown, different weights (both *d* and *e* weights) are represented by different colours, although the *d* values and the *e* values might be different. All graphs are transitive and all except panel e are bi-transitive.



Box 1 Figure 1 | Examples of transitive graphs. **a**, A 'complete' graph. **b**, A cycle. Nodes are joined only to immediate neighbours. **c**, An island structure[16]. Here there are three demes of size 4. Within each deme, all edges have equal weight; edges between demes have another weight. **d**, Two pentagonal cycles with constant edge weights. Edges between cycles have another weight. This is a version of the Petersen graph. **e**, This graph can be thought of in different ways, but we have drawn it to emphasize its relation to the Petersen graph in **d**. In this case there are two cycles of size 3 (triangles), and each node is joined to two nodes in the other triangle with edges of different weights. This graph is symmetric and transitive, but when the blue and amber edges have different *d* or *e* weights it is not bi-transitive. If one takes two nodes in different triangles, one can find an isomorphism that interchanges them, but this is not true for two nodes in the same triangle. Nevertheless, it turns out that equations (1) and (2) do hold for this graph.

Box 2 | Frank's island model of competition[16,18,19]

In more general phenotypic models, fecundity is expanded in a Taylor series:

$$F(X,Y)=F_0+\frac{\partial F}{\partial X}\mathrm{d}X+\frac{\partial F}{\partial Y}\mathrm{d}Y$$

where we take the differential phenotype to be proportional to genotype x (frequency of A): $\mathrm{d}X=\delta x$ where δ is small. This allows us to ignore higher order terms in the phenotypic change and we get weak selection and additivity all at once. In this case, our equations (1) and (2) apply with $c=-\delta\frac{\partial F}{\partial X}$ and $b=\delta\frac{\partial F}{\partial Y}$. In Frank's island model (Box 1 Fig. 1d with n demes of size m, so that population size is $N=nm$) the reproductive resources, $S=S(Z)$, available to a deme decrease with average deme phenotype Z, whereas an individual's share of the deme's resources is proportional to X/Z, its relative competitiveness in the deme; thus $F(X,Y)=\frac{X}{Z}S(Z)$. We take an individual's interactants to be its deme mates, so that $mZ=X+(m-1)Y$. With a BD protocol, equation (1) easily shows that, no matter what the pattern of offspring dispersal, the level of competitiveness X_0 will increase provided

$$\mathrm{BD}:\ \frac{X_0(-S'(X_0))}{S(X_0)}<\frac{(m-1)n}{n-1} \qquad (4)$$

For the DB protocol, we need to specify the dispersal pattern. If h is the probability an offspring remains on its natal deme, then equation (2) shows that X_0 will increase provided

$$\mathrm{DB}:\ \frac{X_0(-S'(X_0))}{S(X_0)}<\frac{(m-1-h)n}{n(1+h)-2} \qquad (5)$$

Provided $h>1/\mathrm{N}$, the right side of equation (4) exceeds that of equation (5), giving a higher evolutionarily stable level of competitiveness in the BD model.

the probability that a single randomly placed A allele in an otherwise B population (B allele in an otherwise A population) will become fixed in the absence of mutation. In the neutral case, $\rho_A=\rho_B=1/N$ as all N individuals have an equal chance of contributing the gene that will ultimately become fixed, and we might measure the advantage of A with the difference $\rho_A-\rho_B$[4,10]. Third, we might use the inclusive fitness effect[12] W_{IF} of the change from B to A (see Methods). To summarize, we might say that A has a selective advantage over B in any of the following situations: $E(A)>E(B)$, $\rho_A>\rho_B$, and $W_{IF}>0$. It is known[1] that for sufficiently small mutation rates in a discrete-generations finite-population model, these three conditions are equivalent to first order in the behavioural deviations b and c. The importance of this for us is that it allows us to use the more easily calculated inclusive fitness effect to measure the direction of change in expected frequency or fixation probability of A. We remark that when b and c are small, the second condition ($\rho_A>\rho_B$) can be shown to be equivalent to $\rho_A>1/N$ in any finite population that can be modelled as a graph (P.D.T. *et al.*, manuscript in preparation). This is not the case[4] for the more general matrix $\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}$ (with a_{hk} small) unless $a_{11}+a_{22}=a_{12}+a_{21}$.

The inclusive fitness effect of an action by a focal individual is defined to be the sum of the fitness effects of this action on all individuals in the population, each effect weighted by its relatedness to the focal individual. To give a simple example, a fecundity gift of b from i to a neighbour j will increase the probability that j will have an offspring but that offspring will replace another neighbour h, so that the inclusive fitness effect will be the extra fecundity to j minus the extra mortality to h, each weighted by the relatedness to i. Of course, if the action has other effects as well, such as a cost c to i, then the effects of that must be accounted for in the same way, and the weak selection assumption (small b and c) will allow us to add these two kinds of effects to obtain the overall inclusive fitness effect. These calculations (see Methods) give us the following simple expressions for the inclusive fitness effect of the allele A in any bi-transitive graph:

$$\text{BD model: } W_{IF}=-b-c(N-1) \qquad (1)$$
$$\text{DB model: } W_{IF}=b[N\bar{d}-2]-c(N-2) \qquad (2)$$

where, in equation (2), $\bar{d}=\Sigma_j e_{ij}d_{ij}$ is the average d-weight of the edge between i and a random interactant, and under transitivity is independent of i. Alternatively, if we pick a random interactant j of i, then $\bar{d}$ is the probability that i's next offspring will displace j.

It is interesting that, in the BD model, the inclusive fitness effect is independent of the structure of the population and of the distribution $\{e_{ij}\}$ of interactants. It is therefore the same as in a random

mixing population in which all d-weights are $d_{ij} = 1/(N-1)$ for $i \neq j$. This generalizes known results[13–15] for an infinite population with an island or stepping-stone structure, and reveals that, under BD, the 'cooperative' allele A can never be selected for positive values of b and c.

In the DB model, the condition does depend on i's interactants, and when these are sufficiently 'close' to i ($\bar{d}$ is large), A can be selectively advantageous for positive b and c with a high enough b/c ratio. The difference between these results for the BD and DB models has to do with where the competitive effects of a 'gift' from i to j are felt. Under BD these are felt by j's neighbours (who, because they include i, have a higher average relatedness to i than does j), whereas under DB these are felt by the neighbours of j's neighbours (who do not, in fact, have a higher average relatedness to i than does j).

These results extend previous results for the case in which selection is both weak and additive. A previous study[10] uses analytical methods to calculate fixation probabilities on the cycle (Box 1 Fig. 1b) for both the BD and the DB models. Our equation (1) is their equation (2.4) and our equation (2), with $\bar{d} = ½$, is their equation (4.4). Their approach works well on the cycle essentially because a population that starts with a single A player will always have the property that the A players are found in a single connected cluster. For more complicated graphs, such as those that involve deme structure (Box 1 Fig. 1c, d), this is not the case and their approach appears to be less tractable. The inclusive fitness analysis continues to hold in these cases as well, however, allowing one to analyse evolution relatively easily in these more complex scenarios provided one has both weak selection and additivity (Box 2).

A previous study[9] used numerical methods and pair approximation techniques to investigate the b–c matrix game in a DB model on a large variety of graphs. The main result of ref. 9 applies to graphs of degree k (every node has k edges with equal weight $d = 1/k$). In their model, partners are always chosen along a connecting edge (that is, $d_{ij} = e_{ij}$ for all ij), and in this case, for $k \ll N$, they find that $\rho_A > 1/N$ when $b/c > k$. They remark that this has the flavour of Hamilton's rule and, for bi-transitive graphs, we now see the connection. Equation (2) tells us that for interactions along all edges (for which $d_{ij} = e_{ij} = 1/k$) the inclusive fitness effect of A is positive when

$$\frac{b}{c} > \frac{k(N-2)}{N-2k} \tag{3}$$

For $k \ll N$ this indeed approximates the ref. 9 condition $b/c > k$. For example, with two connected cycles, each of size 18, and with $k = 3$, equation (3) gives $b/c > 3.4$, which is very nearly $b/c > k$. If the cycles are only of size 5 however (as in the Peterson graph, Box 1 Fig. 1d), then we no longer have $k \ll N$ and equation (3) gives $b/c > 6$. Figure 1 presents simulation results illustrating that equation (3) accurately predicts the spread of a cooperative allele across a range of values of N, even with moderate fitness effects.

Equations (1) and (2) reveal that the ability of cooperation to evolve depends critically on the nature of population regulation. For some forms of population regulation (the BD protocol) equation (1) demonstrates that population viscosity should not affect the evolution of cooperation under a much broader range of conditions than previously appreciated. At the same time, however, equation (2) demonstrates that viscosity under other forms of population regulation (the DB protocol) can facilitate the evolution of cooperation. This inequality clearly delineates the conditions under which this will occur for a quite general class of population structures. It shows that in a sufficiently 'homogeneous' population, if viscosity acts so that interactants have a high likelihood of displacing one another when they reproduce, then cooperation can evolve regardless of the precise form of population structure. Furthermore, although our main focus has been on cooperation, these results apply to the evolution of any trait. For example, in Box 2 we analyse Frank's island model of competition.

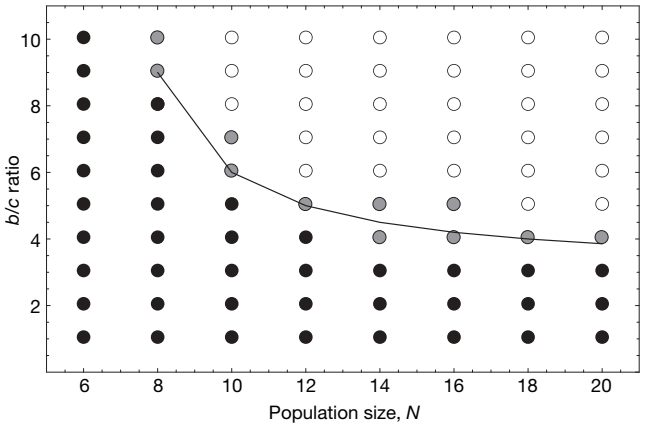


Figure 1 | Simulation results for the DB model with bi-transitive graphs of degree 3, and population sizes $N = 6$ to $N = 20$. The solid line is an analytical prediction from equation (3), above which the cooperative allele is predicted to invade and below which it is not. The population was initialized with one A allele at a randomly chosen node, and the DB process was simulated until fixation of either A or B occurred. Individual fecundities were calculated as $f_i = 1 - cI_i + \sum_j e_{ij} I_j b$, where I_j is an indicator variable equalling 1 if the allele at node j is A, and 0 otherwise. The b/c ratio was varied from 1 to 10, with $c = 0.1$ in all simulations. For each circle, 10,000 realizations of the simulation were run. The 95% confidence interval for the probability of fixation of A was then calculated as $\hat{\rho} \pm \sqrt{\hat{\rho}(1-\hat{\rho})/N}$ where $\hat{\rho}$ = the number of fixations/10^4. If the confidence interval contained the neutral fixation probability, $1/N$, the circle was coloured grey (not different from neutral). If the confidence interval lay entirely below $1/N$ the circle was coloured black (no invasion). If the confidence interval lay entirely above $1/N$ the circle was coloured white (invasion). Similar results are obtained if we instead compare the fixation probability of allele A with that of allele B (rather than comparing it to $1/N$; T.D., unpublished data).

Considerable attention has been paid to the comparison of fixation probabilities on graphs under different matrix games. Such probabilities can be readily calculated only for a restricted class of simple graphs in which each state (number of A alleles) occurs in only one population configuration (for example, regular graphs and cycles). For more general scenarios, simulations and pair approximations have been used instead[9]. In this study, we use a different set of approximations (gene action is weak and additive), and demonstrate that inclusive fitness calculations can predict relative fixation probability in a large class of graphs. In particular, for graphs with a particular homogeneity property, and for small b and c, allele A will have a higher fixation probability than B ($\rho_A > \rho_B$) when the inclusive fitness effect W_{IF} in equations (1) and (2) is positive.

METHODS

Details of the methods used are found in Supplementary Information. The inclusive fitness calculation has two technical components: the calculation of relatedness[17] and the calculation of the inclusive fitness effect of the A allele. For the first, we let G_{ij} be the coefficient of identity-by-descent between nodes i and j, and we find the expected change in these coefficients owing to a single breeding event (which displaces either i or j). The equilibrium condition is obtained by setting this change to be zero. We simplify this condition with the observation that bi-transitivity implies that the matrices $[d_{ij}]$ and $[G_{ij}]$ commute, and we get from this the two key relationships needed in the inclusive fitness analysis:

$$\sum_j d_{ji} G_{jk} = G_{ik} + \mu \ (i \neq k)$$

$$\sum_j G_{jk} d_{jk} = 1 - \mu(N-1)$$

valid to first order in the mutation rate μ.

The inclusive fitness calculation begins with a uniform B population and replaces B with A in a focal actor i. The inclusive fitness effect is the weighted sum of the effects of this replacement on the fitness of each individual j, where the weights are the relatedness between i and j, and we show that these can be taken to be the coefficients G_{ij}. In these calculations, we must keep in mind that any

change in fecundity of an individual will affect both the birth rate and the death rate of neighbouring individuals. The precise effect on these individuals will differ between our two protocols, BD and DB.

Received 23 January; accepted 29 March 2007.

1. Rousset, F. & Billiard, S. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J. Evol. Biol.* **13,** 814–825 (2000).
2. Taylor, P. D., Irwin, A. J. & Day, T. Inclusive fitness in finite deme-structured and stepping-stone populations. *Selection* **1,** 83–93 (2000).
3. Proulx, S. R. & Day, T. What can invasion analyses tell us about evolution under stochasticity? *Selection* **2,** 1–16 (2001).
4. Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428,** 646–650 (2004).
5. Wild, G. & Taylor, P. D. Fitness and evolutionary stability in game theoretic models of finite populations. *Proc. R. Soc. Lond. B* **271,** 2345–2349 (2004).
6. Lessard, S. Long-term stability from fixation probabilities in finite populations: new perspectives for ESS theory. *Theor. Popul. Biol.* **68,** 19–27 (2005).
7. Orzack, S. H. & Hines, W. G. S. The evolution of strategy variation: will an ESS evolve? *Evolution* **59,** 1183–1193 (2005).
8. Lieberman, E., Hauert, C. & Nowak, M. A. Evolutionary dynamics on graphs. *Nature* **433,** 312–316 (2005).
9. Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. A. A simple rule for the evolution of cooperation on graphs. *Nature* **441,** 502–505 (2006).
10. Ohtsuki, H. & Nowak, M. A. Evolutionary games on cycles. *Proc. R. Soc. B* **273,** 2249–2256 (2006).
11. Moran, P. A. P. *Statistical Processes of Evolutionary Theory* (Oxford, Clarendon, 1962).
12. Hamilton, W. D. The genetical evolution of social behaviour, I and II. *J. Theor. Biol.* **7,** 1–52 (1964).
13. Wilson, D. S., Pollock, G. B. & Dugatkin, L. A. Can altruism evolve in a purely viscous population? *Evol. Ecol.* **6,** 331–341 (1992).
14. Taylor, P. D. Altruism in viscous populations – an inclusive fitness model. *Evol. Ecol.* **6,** 352–356 (1992).
15. Taylor, P. D. Inclusive fitness in a homogeneous environment. *Proc. R. Soc. Lond. B* **249,** 299–302 (1992).
16. Wright, S. Isolation by distance. *Genetics* **28,** 114–138 (1943).
17. Michod, R. E. & Hamilton, W. D. Coefficients of relatedness in sociobiology. *Nature* **288,** 694–697 (1980).
18. Frank, S. A. Kin selection and virulence in the evolution of protocells and parasites. *Proc. R. Soc. Lond. B* **258,** 153–161 (1994).
19. Taylor, P. D. & Frank, S. How to make a kin selection argument. *J. Theor. Biol.* **180,** 27–37 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements We thank D. Gregory for an exchange of ideas, and A. Gardner, J. Pepper and A. Grafen for many comments. This research was funding by grants to P.D.T. and T.D. from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Author Contributions All authors contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.D.T. (peter.taylor@queensu.ca).

LETTERS

An autopodial-like pattern of Hox expression in the fins of a basal actinopterygian fish

Marcus C. Davis[1], Randall D. Dahn[1] & Neil H. Shubin[1,2]

Comparative analyses of Hox gene expression and regulation in teleost fish and tetrapods support the long-entrenched notion that the distal region of tetrapod limbs, containing the wrist, ankle and digits, is an evolutionary novelty[1–4]. Data from fossils support the notion that the unique features of tetrapod limbs were assembled over evolutionary time in the paired fins of fish[5]. The challenge in linking developmental and palaeontological approaches has been that developmental data for fins and limbs compare only highly derived teleosts and tetrapods; what is lacking are data from extant taxa that retain greater portions of the fin skeletal morphology considered primitive to all bony fish[6,7]. Here, we report on the expression and function of genes implicated in the origin of the autopod in a basal actinopterygian, *Polyodon spathula*. *Polyodon* exhibits a late-phase, inverted collinear expression of 5′ HoxD genes, a pattern of expression long considered a developmental hallmark of the autopod and shown in tetrapods to be controlled by a 'digit enhancer' region. These data show that aspects of the development of the autopod are primitive to tetrapods and that the origin of digits entailed the redeployment of ancient patterns of gene activity.

The acquisition of hands and feet (autopodia) was a seminal event in tetrapod evolution, facilitating terrestrial invasion and habitation. Digits—elongate, segmented rods with individuated morphologies—have no clear antecedents in extant fish fins, and so the autopod has generally been regarded as a tetrapod innovation. During early stages of teleost and tetrapod appendage development, 5′ members of the homeobox A (HoxA) and HoxD gene clusters are expressed in a conserved, spatiotemporally collinear fashion; more 5′ genes are progressively activated, and expressed in domains progressively restricted along the proximodistal and anteroposterior axes, respectively. However, at later stages when autopodial elements are being determined and patterned, a distinct tetrapod-specific 'late phase' of 5′ HoxD expression is activated with inverted spatial collinearity along the anteroposterior axis. Concomitantly, *Hoxa11* and *Hoxa13* resolve into mutually exclusive domains, with *Hoxa13* expressed throughout the autopod. Late-phase Hox expression is not observed in teleosts, and it has been proposed that these regulatory changes in Hox expression may underlie the origin of the autopod[1,2]. Indeed, a DNA regulatory element recently identified in mouse controls late-phase autopodial Hox expression, and has been dubbed the 'digit enhancer'; this element has not been found in teleosts[3]. An alternative hypothesis is that teleosts may have lost or modified portions of an ancestral Hox program that is retained in tetrapods. Morphological data lends support to this notion: teleosts appear to have lost skeletal structures in the fin that are homologous to tetrapod limb bones[6,7].

To address these issues, we have analysed genes of the *Sonic hedgehog* (*Shh*) pathway and HoxA and HoxD cluster genes in a basal actinopterygian, the paddlefish *Polyodon spathula*. *Polyodon* possesses pectoral fin endoskeletal elements considered homologous to both teleosts radials and tetrapod limb bones[6,7] (see also Supplementary Information). Furthermore, analysis of fin development in *Polyodon* reveals a mosaic pattern of endoskeletal condensation and chondrogenesis, exhibiting aspects of both teleost and tetrapod appendage development (ref. 6 and Supplementary Information). For example, the posterior portion of the fin, the metapterygium, contains branched radials comparable to sarcopterygians such as *Eusthenopteron* and *Tiktaalik*[5]. In addition, the boundary between proximal and distal radials is comparable to the proximal portion of the autopod. In both taxa, a boundary between rod-like endochondral bones and more distal nodular bones transverses the width of the appendage. All of these features are general to many vertebrates, so it may well be that the development of these regions is reflected in common genetic patterns among the tetrapod autopod and different regions of fish fins.

As is expected from analysis of *Shh* expression in teleosts[1] and chondrichthyans[8], major components of the *Shh* pathway in limbs are conserved in *Polyodon*. *Shh* expression is restricted to a posterior domain comparable to the zone of polarizing activity (ZPA) of model taxa (Fig. 1a). Exposure to retinoic acid (RA) before fin formation (stages 37–39) results in a broadened anterior expression of *Shh* (Fig. 1e, see also ref. 8) comparable to RA-affected zebrafish[9], but differing from the focused 'ectopic ZPA' described in tetrapods[10]. As was predicted, exposure to a cyclopamine analogue (SANT-1) inhibited or downregulated *Shh* expression below detectable levels (Fig. 1f) and buds failed to develop a fin skeleton (data not shown). We also assessed expression of the *Shh*-associated transmembrane protein *Ptc1*. Wild-type expression of *Ptc1* is similar to that of *Shh* (Fig. 1b). *Ptc1* expression in RA-affected fins is broadened anteriorly, as is *Shh* (Fig. 1g). In SANT-1 fins, *Ptc1* expression is similarly inhibited or downregulated (Fig. 1h). To support further the interpretation that the appendage *Shh* pathway in *Polyodon* is conserved, we looked at the expression of the transcription factor *dHAND*, thought to be an upstream activator of *Shh*[11], and *Gli3*, which is considered a direct intracellular mediator of *Shh*[12]. *dHAND* is expressed posteriorly in the mouse limb before *Shh*, but in a broader domain than that of *Shh*/ZPA[11], and at later stages is expressed in the interdigital tissue[13]. Mutations in the Hand2 locus in zebrafish lack anteroposterior patterning in the fin[14]. In *Polyodon*, *dHAND* is expressed early in the posterior mesoderm, but then expands anteriorly across the distal bud (Fig. 1c, stages 42 and 43). As in tetrapods[13], late-stage expression of *dHAND* in *Polyodon* is restricted to the distal-most cells of the bud, in this case to the cells that will form the distal radials (Fig. 1c, stage 46). *Gli3* is expressed throughout the early mouse limb before 5′ Hox-mediated expression of *Shh* restricts *Gli3* to the anterior bud[15]. In *Polyodon*, *Gli3* is initially expressed throughout the fin bud, and later becomes restricted to distal cells in a similar manner to *dHAND* (Fig. 1d).

[1]Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, Illinois 60637, USA. [2]Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, Illinois 60605, USA.

Early HoxA expression is remarkably similar in tetrapods and zebrafish. In both, *Hoxa11* and *Hoxa13* exhibit a proximodistally nested collinear expression, with *Hoxa13* the most distally restricted[1,2,16]. Later expression is markedly different: during the stages of digit specification, tetrapod *Hoxa11* becomes restricted to the zeugopod region and *Hoxa13* becomes restricted to the developing autopod[1,2]. In zebrafish, *Hoxa11* and *Hoxa13* remain nested and overlapping with no observed proximodistal segregation[1,16].

We looked at *Hoxa11* and *Hoxa13* expression in *Polyodon* from onset to cessation of expression to ascertain which of these conditions might be primitive. Early HoxA expression is similar to that of tetrapods and zebrafish (Fig. 2a, b; stage 40). *Hoxa11* is expressed through the posterior two-thirds of the fin bud and is also expressed proximally in the dorsal and ventral myoblasts (white arrowhead in Fig. 2a; see also ref. 17). *Hoxa13* expression is restricted to the distal-most fin bud and is never observed proximally. In subsequent stages, *Hoxa11* and *Hoxa13* become progressively more distally restricted (Fig. 2a, b; stages 44 and 46) and disappear after stage 46. Expression remains nested and overlapping throughout development, with no evidence of tetrapod-like proximodistal segregation.

Tetrapods and zebrafish exhibit a conserved early phase of HoxD expression, in which transcripts of the more 5′ genes are expressed in a posteriorly restricted nested pattern[1,18], that may have evolved in the unpaired fins of early vertebrates[19]. In tetrapods, a late phase of HoxD expression extends anteriorly across the distal portion of the limb that will form the autopod[4,18]. Zebrafish appear to lack this late phase of HoxD and also lack elaborate distal skeletal structures resembling digits. On the basis of these observations, it has been hypothesized that this late phase of HoxD expression is a tetrapod novelty correlated with the evolution of the autopod[1,16].

To test this hypothesis, we looked at expression of 5′ HoxD genes in *Polyodon*. Early-phase expression of *Hoxd11–13* exhibits the same posteriorly nested pattern seen in other vertebrates, with *Hoxd13* being the most anteroposteriorly restricted (Fig. 3a–c; stage 40).

Remarkably, at later stages there is a distally restricted and inverted collinear phase of 5′ HoxD expression in the fins of *Polyodon* (Fig. 3a–c; stage 46). At this stage, *Hoxd13* expression extends more anteriorly than either *Hoxd11* or *Hoxd12*. Later still (Fig. 3a–c; at ten days post-staging, 10 d.p.s.), 5′ HoxD expression expands proximally into the region surrounding, but not including, the developing radials (Fig. 3g–i). *Hoxd11* and *Hoxd12* expression extends anteriorly to the inter-radial space between mesopterygial radials one and two (Fig. 3a–c; compare to Fig. 1k), while *Hoxd13* expression extends all the way to the anterior margin of the radial field, demarking the space between mesopterygial radial 1 and the propterygium and the space anterior to the propterygium (Fig. 3c, 10 d.p.s.). Exposure to RA results in ectopic anterior expansion of early-phase *Hoxd11* (Fig. 3d). Late-phase *Hoxd11* expression in RA-exposed embryos is expanded both anteriorly and proximally (Fig. 3e; compare with stage 46 in Fig. 3a), in agreement with the observed *Shh* responsiveness of late-phase HoxD in tetrapods[18]. Also consistent with results from tetrapods[20], SANT-1-mediated inhibition of *Shh* does not affect early-phase *Hoxd11* expression (Fig. 3f).

Polyodon shows a mosaic of primitive and derived features of skeletal pattern, appendage development, and gene activity. All bony fish surveyed to date share a conserved pattern of expression of *Shh* pathway, early-phase HoxA and HoxD gene expression, and appendage skeletons that can be generalized to an early developing proximal region and a later developing distal region. However, *Polyodon* and

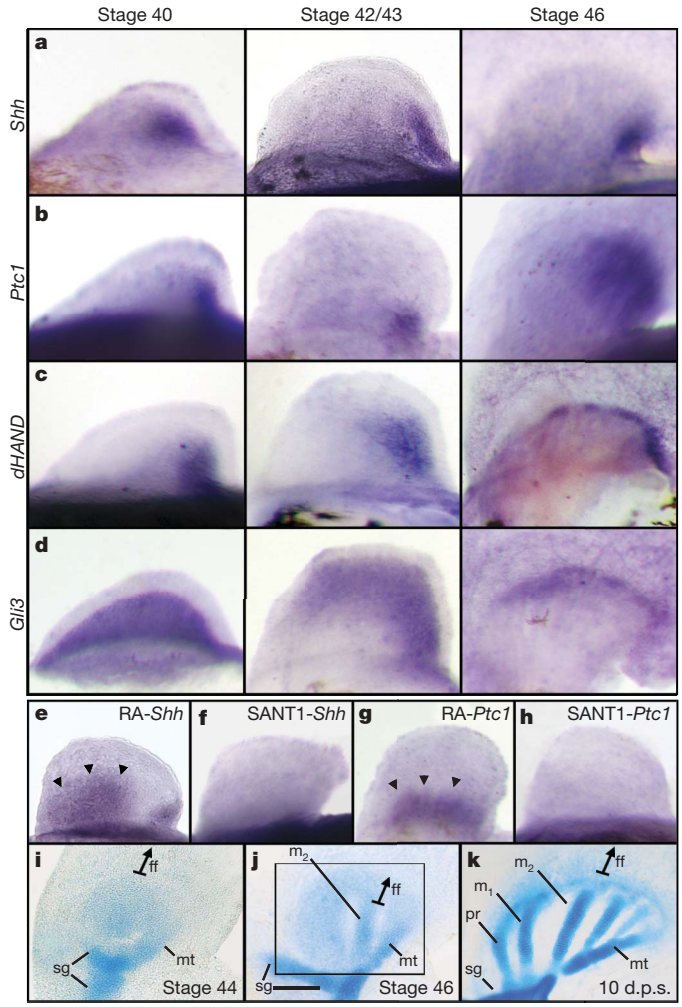


Figure 1 | Expression of *Shh* pathway genes in *Polyodon* pectoral fins by stage. Anterior to left; distal to top. **a**–**d**, *Shh* (**a**), *Ptc1* (**b**), *dHAND* (**c**) and *Gli3* (**d**). *Shh* and *Ptc1* remain in the posterior fin mesenchyme throughout development. *dHAND* expression is initially posterior (stage 40) but then becomes distally restricted (stage 46). *Gli3* is initially expressed throughout the fin mesenchyme (stage 40) before following a similar pattern of expression to *dHAND*. **e**, **f**, Exposure to RA (**e**) results in ectopic expression of *Shh* in the anterior two-thirds of the fin (arrowheads), while SANT-1 (**f**) downregulates normal *Shh* expression. **g**, **h**, Expression of *Ptc1* in animals exposed to RA (**g**) and SANT-1 (**h**) show similar ectopic expression (arrowheads) and downregulation respectively. **i**–**k**, Alcian Blue skeletal preparations of late larval stages (**i**, **j**) and a fry at 10 d.p.s. (**k**). See Supplementary Information for a more detailed series. Boxed region in **j** corresponds to stage 46 figures in **a**–**d**. ff, fin fold; m_1, mesopterygial radial 1; m_2, mesopterygial radial 2; mt, metapterygium; pr, propterygium; sg, shoulder girdle.

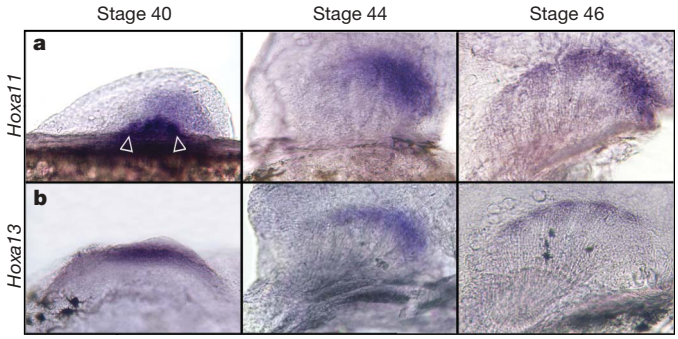


Figure 2 | Expression of HoxA genes in *Polyodon* pectoral fins. Anterior to left; distal to top. **a**, **b**, *Hoxa11* (**a**), and *Hoxa13* (**b**). At early stages (**a**, stage 40), *Hoxa11* is expressed throughout the posterior mesenchyme and strongly expressed in the nascent dorsal and ventral muscle buds (open arrowheads). *Hoxa11* then becomes increasingly restricted to the distal fin bud (**a**, stage 44) and disappears completely shortly after the onset of feeding (**a**, stage 46). *Hoxa13* is restricted to the distal-most fin bud cells from onset of expression (**b**, stage 40) until expression disappears (**b**, stage 46). *Hoxa11* and *Hoxa13* remain nested and overlapping throughout development. Compare stages 44–46 to skeletal development in Fig. 1i and j respectively.

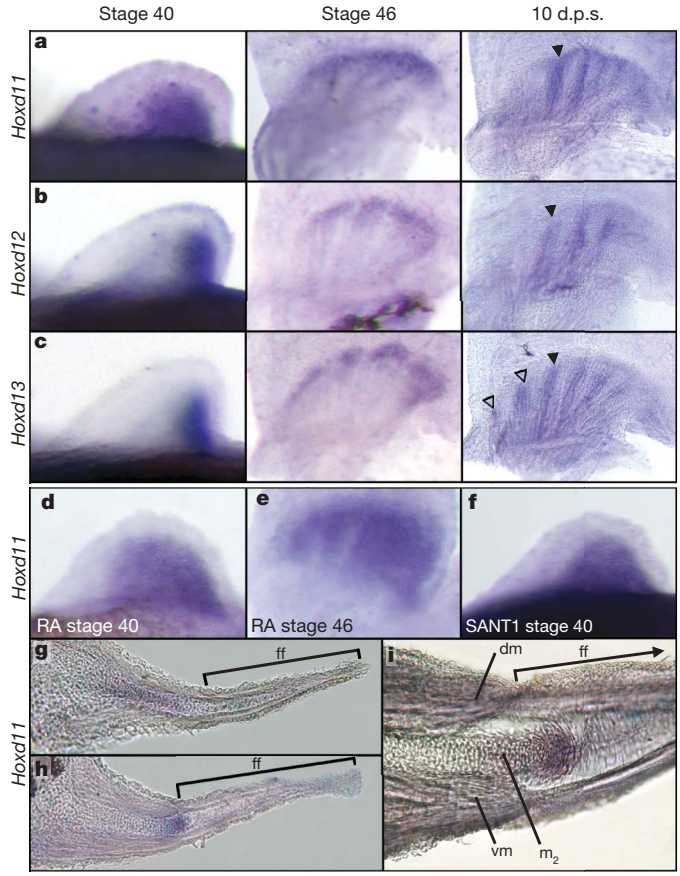


Figure 3 | Expression of HoxD genes in the pectoral fin of *Polyodon*. Anterior to the left; distal to top in **a**–**f**. Proximal to the left; dorsal to the top in **g**–**i**. **a**–**c**, Expression of *Hoxd11* (**a**), *Hoxd12* (**b**) and *Hoxd13* (**c**). Early HoxD expression (stage 40) exhibits the collinear pattern described for other vertebrates. Late-phase HoxD expression is initially restricted to the cells surrounding condensing radials (stage 46) and the nascent distal radials, but then extends proximally to the inter-radial cells (10 d.p.s.). The anterior limit of *Hoxd11* and *Hoxd12* expression is between m1 and m2 (closed arrowheads). In contrast, *Hoxd13* is expressed more anteriorly (open arrowheads). **d**–**f**, Expression of *Hoxd11* in animals exposed to RA (**d**, **e**) and SANT-1 (**f**). RA slightly upregulates early-phase *Hoxd11* expression (**d**, compare with stage 40 in **a**) and broadly upregulates late-phase *Hoxd11* expression (**e**, compare with stage 46 in **a**). SANT-1 has no effect on *Hoxd11* expression (**f**, compare with stage 40 in **a**). **g**–**i**, Transverse sections through 10 d.p.s. pectoral fins expressing *Hoxd11*. Expression is restricted to inter-radial (**g**, section at the anterior–posterior level of closed arrowhead in **a**) and in adjacent distal-most cells (**h**, section posterior to the level of the closed arrowhead in **a**). Magnification (**i**) reveals that *Hoxd11* expression is isolated to the mesenchymal cells distal to the proximal radials. This region sits within the proximal fin fold and is the site of condensing distal radials (see Supplementary Information). dm, dorsal fin musculature; vm, ventral musculature.

tetrapods share a developmental hallmark that is not observed in teleost fins: an inverted collinear expression of HoxD genes in the distal region of the appendage. Novelty in the appendages of lobe-finned and ray-finned fish has arisen both by changes in regulation, as in tetrapods, and by loss of portions of an ancient and conserved pattern of Hox expression in teleosts.

METHODS SUMMARY

Embryos and staging. *Polyodon spathula* eggs were incubated at 18 °C (12 h:12 h light:dark cycle) in 25% Hank's Balanced Salt Solution (HBSS) until hatching, and grown thereafter at 22 °C. Embryos were staged as described[6]. Alcian Blue cartilage preparations and expression analyses were performed as previously described, respectively[6,21].

Polyodon embryos ($n = 100$) were treated with all-trans retinoic acid (1×10^{-6}; Sigma R2625) in 25% HBSS + 0.5% dimethylsulphoxide (DMSO) for 30 min in the dark at stage 37 (pre-fin budding) or stage 39 (pectoral fin endoskeletal condensation). Embryos were returned to 25% HBSS after several washes to remove residual retinoic acid. Embryos exposed to the *Hedgehog* inhibitor SANT-1 (10 μM in 25% HBSS + 0.5% DMSO; Calbiochem 559303) were maintained in this solution until harvesting. Control embryos ($n = 100$; 25% HBSS containing 0.5% DMSO) showed no differences from wild-type expression or skeletal pattern. Mortality rates were approximately 50% for both RA and SANT-1 exposure.

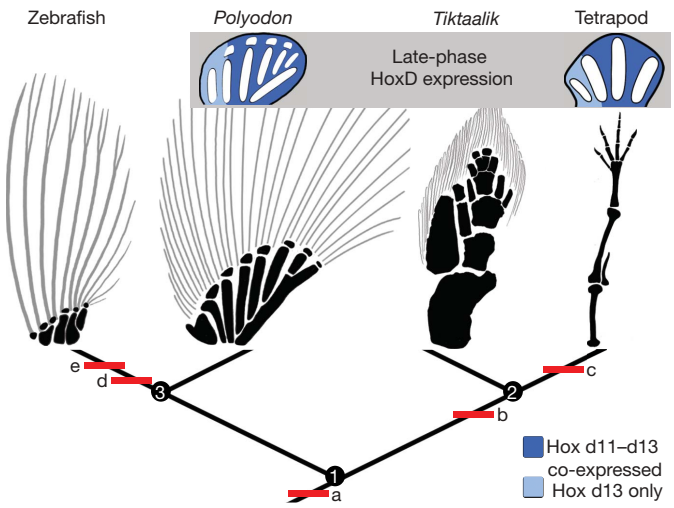


Figure 4 | Conservation of late-phase HoxD expression in bony fish (osteichthyes). Anterior to the left; distal to top. HoxD expression in *Polyodon* supports the notion that late-phase HoxD expression is primitive to tetrapods and to osteichthyes in general. Red bars denote: **a**, acquisition of late-phase HoxD expression; **b**, loss of non-metapterygial radials; **c**, loss of the dermal fin fold; **d**, loss of the metapterygium; **e**, loss of late-phase HoxD expression. Blue cartoons depict conserved late-phase HoxD seen in *Polyodon* and tetrapods (here represented by the hind limb of the chicken, *Gallus*).

Isolation of *Polyodon* genes. *Polyodon* gene fragments were isolated by RT–PCR using the Expand High Fidelity PCR System (Roche) and manufacturer's instructions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 February; accepted 13 April 2007.

1. Sordino, P., van der Hoeven, F. & Duboule, D. Hox gene expression in teleost fins and the origin of vertebrate digits. *Nature* **375**, 678–681 (1995).
2. Wagner, G. P. & Chiu, C.-H. The tetrapod limb: A hypothesis on its origin. *J. Exp. Zool.* **291**, 226–240 (2001).
3. Spitz, F., Gonzalez, F. & Duboule, D. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405–417 (2003).
4. Shubin, N., Tabin, C. & Carroll, S. Fossils, genes, and the evolution of animal limbs. *Nature* **388**, 639–648 (1997).
5. Shubin, N. H., Daeschler, E. B. & Jenkins, F. A., Jr. The pectoral fin of *Tiktaalik roseae* and the origin of the tetrapod limb. *Nature* **440**, 747–749 (2006).
6. Davis, M. C., Shubin, N. H. & Force, A. Pectoral fin and girdle development in the basal actinopterygians *Polyodon spathula* and *Acipenser transmontanus*. *J. Morphol.* **262**, 608–628 (2004).
7. Mabee, P. M. Developmental data and phylogenetic systematics: evolution of the vertebrate limb. *Am. Zool.* **40**, 789–800 (2000).
8. Dahn, R. D., Davis, M. C., Pappano, W. N. & Shubin, N. H. Sonic hedgehog function in chondrichthyan fins and the evolution of appendage patterning. *Nature* **445**, 311–314 (2007).
9. Hoffman, L., Miles, J., Avaron, F., Laforest, L. & Akimenko, M.-A. Exogenous retinoic acid induces a stage-specific, transient and progressive extension of *Sonic hedgehog* expression across the pectoral fin bud of zebrafish. *Int. J. Dev. Biol.* **46**, 949–956 (2002).
10. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. *Sonic hedgehog* mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).
11. Charité, J., McFadden, D. G. & Olson, E. N. The bHLH transcription factor *dHAND* controls *Sonic hedgehog* expression and establishment of the zone of polarizing activity during limb development. *Development* **127**, 2461–2470 (2000).

12. Altaba, A. R. Gli proteins encode context-dependent positive and negative functions: implications for development and disease. *Development* **126,** 3205–3216 (1999).
13. Fernandez-Teran, M. *et al.* Role of dHAND in the anterior-posterior polarization of the limb bud: Implications for the Sonic hedgehog pathway. *Development* **127,** 2133–2142 (2000).
14. Yelon, D. *et al.* The bHLH transcription factor Hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development* **127,** 2573–2582 (2000).
15. Litingtung, Y., Dahn, R. D., Li, Y., Fallon, J. F. & Chiang, C. Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity. *Nature* **418,** 979–983 (2002).
16. Sordino, P. & Duboule, D. A molecular approach to the evolution of vertebrate paired appendages. *Trends Ecol. Evol.* **11,** 114–119 (1996).
17. Metscher, B. D. *et al.* Expression of Hoxa-11 and Hoxa-13 in the pectoral fin of a basal ray-finned fish, *Polyodon spathula*: implications for the origin of tetrapod limbs. *Evol. Dev.* **7,** 186–195 (2005).
18. Tarchini, B. & Duboule, D. Control of Hoxd genes' collinearity during early limb development. *Dev. Cell* **10,** 93–103 (2006).
19. Frietas, R., Zhang, G. & Cohn, M. J. Evidence that mechanisms of fin development evolved in the midline of early vertebrates. *Nature* **442,** 1033–1037 (2006).
20. Ros, M. A. *et al.* The chick *oligozeugodactyly* (*ozd*) mutant lacks *sonic hedgehog* function in the limb. *Development* **130,** 527–537 (2003).
21. Prince, V. E., Joly, L., Ekker, M. & Ho, R. K. Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. *Development* **125,** 407–420 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Kahrs family and Osage Catfisheries for their continuing support of our research. This project is dedicated to the memory of John and Jim Kahrs, without whom *Polyodon* would not have been available for this research. M.C.D is supported by the University of Chicago, N.H.S. is supported by grants from the University of Chicago, the NSF and the NIH (NRSA to R.D.D.).

Author Information Sequences for *Polyodon Shh, Ptc1, dHAND, Gli3, Hoxa11, Hoxa13, Hoxd11, Hoxd12,* and *Hoxd13* are deposited in Genbank with accession numbers: *Shh* (1,006 bp, EF100659), *Ptc1* (1,253 bp, EF527815), *dHAND* (278 bp, EF527816), *Gli3* (1,045 bp, EF534083), *Hoxd11* (486 bp, EF527819), *Hoxd12* (420 bp, EF527820), and *Hoxd13* (863 bp, EF527821), *Hoxa11* (660 bp, EF527817), *Hoxa13* (833 bp, EF52718). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.H.S. (nshubin@uchicago.edu).

METHODS

Embryos and staging. Fertilized *Polyodon spathula* eggs were acquired at four days post-fertilization from Osage Catfisheries (Osage beach, Missouri). Larvae were reared (22 °C, 12 h:12 h light:dark cycle) in 25% HBSS and fed *Artemia* from the onset of feeding behaviour (stage 46) until harvesting. Embryos and larvae were staged as previously described[6,22]. *Polyodon* were euthanized with a lethal dose of MS-222 (tricaine) and fixed for 24 h in 4% paraformaldehyde, dehydrated step-wise into 100% methanol, and stored until use at −20 °C.

Isolation of *Polyodon* genes. Amplified *Polyodon* gene fragments were cloned into PGEM-T Easy Vector (Promega). Primers were designed against conserved regions within the homeobox of *HoxA* and *HoxD* genes and against conserved 5′ and 3′ regions of Shh-pathway-related genes for mouse (*Mus*), chicken (*Gallus*), zebrafish (*Danio*), and horned shark (*Heterodontus)* available on GenBank.

Whole-mount *in situ* hybridization. These were performed as described[21] with the following modifications: All washes took place in 1.5 ml Eppendorf tubes with liquid decanted each wash; an additional PBT (phosphate buffered saline + 0.2% Tween-20) wash day (eight washes) was added between anti-digoxigenin incubation and staining; and BM Purple (Roche Labs) was used to develop the signal instead of nitro blue tetrazolium + 5-bromo-4-chloro-3-indolyl phosphate. Specimens were cleared step-wise into 100% glycerol for imaging.

Skeletal preparations. Alcian Blue skeletal preparations were performed as previously described[6].

Histology. Thick sections were prepared by hydrating whole-mount *in situ* stained specimens in phosphate-buffered saline and embedding in 17% gelatin (dissolved in 10% HBSS at 37 °C). Specimens were then fixed for 24 h in 4% paraformaldehyde, washed in phosphate-buffered saline, and mounted for sectioning. Sections were cut to 30 μm thickness on a Lancer 1000 vibratome, mounted on glass slides, coated with 75% glycerol, and photographed.

22. Bemis, W. E. & Grande, L. Early development of the actinopterygian head. I. External development and staging of the paddlefish *Polyodon spathula*. *J. Morphol.* **213**, 47–83 (1992).

LETTERS

Transcriptional coactivator PGC-1α integrates the mammalian clock and energy metabolism

Chang Liu[1], Siming Li[1], Tiecheng Liu[2], Jimo Borjigin[2] & Jiandie D. Lin[1]

The mammalian clock regulates major aspects of energy metabolism, including glucose and lipid homeostasis and mitochondrial oxidative metabolism[1,2]. The biochemical basis for coordinated control of the circadian clock and diverse metabolic pathways is not well understood. Here we show that PGC-1α (Ppargc1a), a transcriptional coactivator that regulates energy metabolism[3–9], is rhythmically expressed in the liver and skeletal muscle of mice. PGC-1α stimulates the expression of clock genes, notably *Bmal1* (*Arntl*) and *Rev-erbα* (*Nr1d1*), through coactivation of the ROR family of orphan nuclear receptors. Mice lacking PGC-1α show abnormal diurnal rhythms of activity, body temperature and metabolic rate. The disruption of physiological rhythms in these animals is correlated with aberrant expression of clock genes and those involved in energy metabolism. Analyses of PGC-1α-deficient fibroblasts and mice with liver-specific knockdown of PGC-1α indicate that it is required for cell-autonomous clock function. We have thus identified PGC-1α as a key component of the circadian oscillator that integrates the mammalian clock and energy metabolism.

Most living organisms exhibit behavioural and physiological rhythms, including activity, sleep, metabolism and body temperature. These rhythms are controlled by a circadian clock and are subject to regulation by light–dark cycles. In mammals, the central and peripheral clocks are controlled by a common transcriptional circuitry that generates rhythmic patterns of gene expression, most notably those encoding metabolic enzymes[10–14]. Perturbations of circadian rhythms have been associated with an increased risk for metabolic disorders in humans[15]. Remarkably, mice with defective clock function develop obesity and have impaired glucose homeostasis[16,17], suggesting that the regulation of circadian clocks is linked to the pathways of energy metabolism and potentially to the pathogenesis of metabolic diseases. PGC-1α is an inducible transcriptional coactivator that regulates adaptive energy metabolism in multiple tissues[4,6]. The critical role of this factor in linking the environment to metabolism prompted us to examine whether PGC-1α regulates the circadian clock and serves as a link between the clock and energy metabolism.

The clock oscillator consists of transcriptional activators and repressors that are assembled into feedback loops[18,19]. To determine the effects of PGC-1α on clock genes, we transduced primary hepatocytes and C2C12 myotubes with recombinant adenoviruses expressing green fluorescent protein (GFP) or PGC-1α. Quantitative PCR (qPCR) analyses indicate that PGC-1α induces the expression of *Bmal1* and *Clock* in both cell types (Fig. 1a). Consistent with messenger RNA expression, Bmal1 and Clock protein levels are elevated in hepatocytes in response to PGC-1α (Fig. 1b). Clock and Bmal1 are two basic helix–loop–helix transcription factors that activate the transcription of period (*Per*) and cryptochrome (*Cry*) genes[18,19]. Per and Cry proteins in turn inhibit their own expression by repressing Clock/Bmal1 activity, forming the critical feedback loop within the clock circuitry. Interestingly, the effects of PGC-1α on *Per* and *Cry* gene expression seem to be cell-type-specific. Previous studies have implicated orphan nuclear receptors of the ROR and Rev-erb families in the control of circadian clock function[13,20–22]. Although PGC-1α exerts only modest effects on the expression of *RORα* (*Rora*) and *RORγ* (*Rorc*), it significantly induces *Rev-erbα* and *Rev-erbβ* expression in both cell types (Fig. 1a).

Recent expression profiling studies have identified a large number of genes that show rhythmic expression in the liver[11–13]. To assess the role of PGC-1α in the regulation of circadian gene expression at the transcriptome level, we performed microarray analysis on total RNA isolated from primary hepatocytes transduced with *GFP* or *PGC-1α* adenoviruses. We integrated the PGC-1α microarray data set with a previously described set of 335 oscillating hepatic genes[11] to determine global effects of PGC-1α on circadian genes. As shown in Fig. 1c, 47 of these circadian-regulated genes are induced by PGC-1α by more than 1.8-fold, with the expression of many clustering at the onset and during early dark phase (Supplementary Table 1). Analysis of *PGC-1α* mRNA levels in mouse liver revealed a diurnal rhythm that peaks at CT13 (CT0 is the onset at hour 0 of subjective light period), gradually declines thereafter, and reaches a nadir at CT1 (Fig. 1d). PGC-1α expression is similarly rhythmic in skeletal muscle, though to a lesser degree. The expression of other PGC-1 coactivators, that is, *PGC-1β* and *PRC*, also show circadian rhythms in these tissues. Immunoblotting analyses indicate that PGC-1α protein reaches peak levels in early dark phase (Fig. 1e), coinciding with the activation of *Bmal1* transcription and its mRNA accumulation. Interestingly, the short isoform of PGC-1β shows a diurnal rhythm that peaks several hours before PGC-1α. In contrast, the protein levels of Brm (also known as *Smarca2*), a component of the SWI/SNF chromatin-remodelling complexes, remain unchanged. Taken together, these data indicate that rhythmic activation of PGC-1α, and possibly other PGC-1 coactivators, may have an important role in the regulation of clock function.

To identify transcription factors that mediate the regulation of clock genes by PGC-1α, we examined the ability of PGC-1α to synergize with these factors in the regulation of *Bmal1* transcription. PGC-1α dramatically augments the transcriptional activity of RORα and RORγ, but not C/EBPα, on a *Bmal1* promoter reporter (Fig. 2a and Supplementary Fig. 1). Coactivation of RORα by PGC-1α was previously observed for genes involved in muscle lipid metabolism[23]. In contrast, PGC-1α exerts modest effects on the transcriptional activity of Clock/Bmal1 (data not shown). The synergistic effects of ROR and PGC-1α are abolished when the ROR-binding sites (RORE) on the proximal *Bmal1* promoter are mutated. Physical interaction assays indicate that PGC-1α binds to RORα and RORγ through the LXXLL motifs (Supplementary Fig. 1). This functional synergy between

[1]Life Sciences Institute and Department of Cell & Developmental Biology, [2]Department of Molecular and Integrative Physiology, University of Michigan Medical Center, Ann Arbor, Michigan 48109, USA.

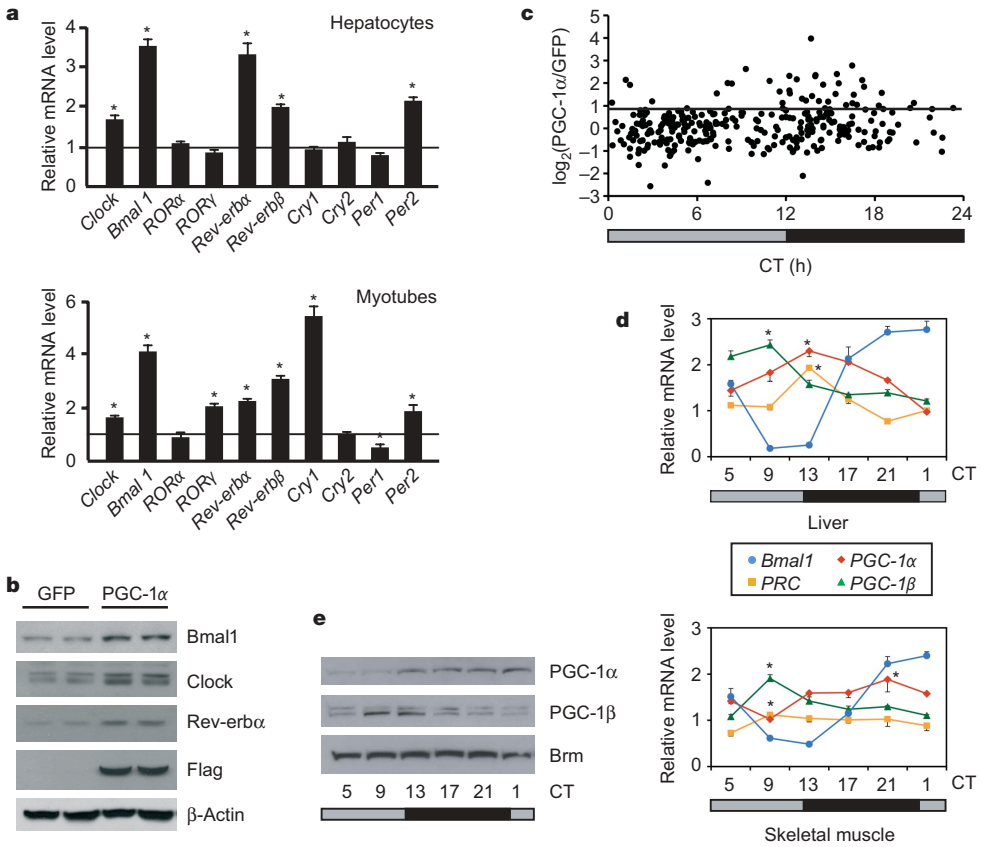


Figure 1 | Regulation of clock genes by PGC-1α. **a**, qPCR analysis of clock genes in primary hepatocytes and C2C12 myotubes. Shown is fold-induction by PGC-1α. Data represent mean ± s.d. $*P < 0.02$, PGC-1α versus GFP. **b**, Immunoblots of hepatocyte lysates using the indicated antibodies. **c**, Effects of PGC-1α on circadian-regulated genes in the liver. Fold induction by PGC-1α of the circadian gene set was calculated on the basis of microarray data, log-transformed and plotted against the respective peak timepoints of expression in the liver (solid line, 1.8-fold). **d**, qPCR analysis of liver and skeletal muscle gene expression. Data represent mean ± s.d. $*P < 0.01$, peak versus nadir. **e**, Immunoblots of liver nuclear extracts using the indicated antibodies.

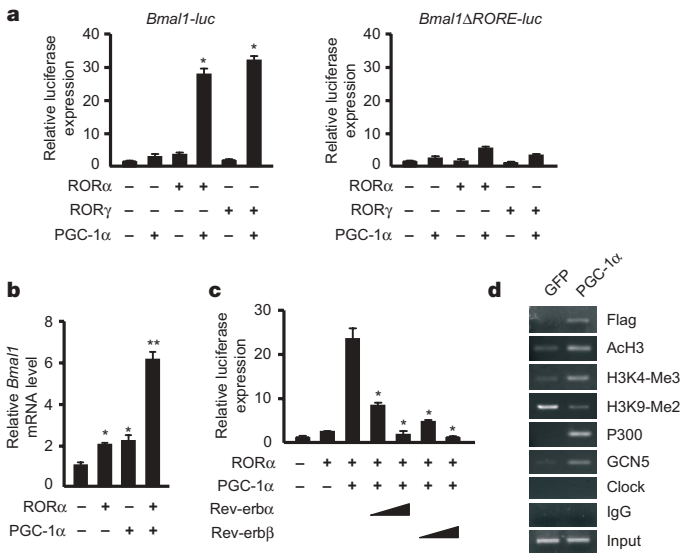


Figure 2 | Regulation of *Bmal1* gene transcription by PGC-1α and ROR. **a**, Reporter gene assays using wild-type (*Bmal1-luc*) or RORE-mutant (*Bmal1ΔRORE-luc*) *Bmal1*-luciferase reporters. $*P < 0.01$, PGC-1α versus control. **b**, qPCR analysis of *Bmal1* expression in transduced hepatocytes. **a**, **b**, Error bars, s.d. $*P < 0.05$, RORα or PGC-1α versus Control; $**P < 0.01$, RORα plus PGC-1α versus RORα. **c**, Reporter gene assays using *Bmal1-luc* in the presence of Rev-erbα or Rev-erbβ. Shown with mean ± s.d. $*P < 0.001$, Rev-erb versus control. **d**, ChIP assay with the indicated antibodies using HepG2 cells transduced with *GFP* or *Flag–PGC-1α* adenoviruses. PCR primers amplify a fragment flanking the proximal RORE.

RORα and PGC-1α is also observed for the endogenous *Bmal1* gene (Fig. 2b). Previous studies indicate that Rev-erbα negatively regulates *Bmal1* transcription by recruiting corepressor proteins[22], indicating that the induction of Rev-erb factors may provide a negative feedback mechanism for PGC-1α function. Consistent with this, both Rev-erbα and Rev-erbβ drastically repress the stimulatory effects of PGC-1α on the *Bmal1-luc* reporter (Fig. 2c). These results illustrate that the ability of PGC-1α to activate *Bmal1* transcription is modulated by the relative abundance of the ROR and Rev-erb families of orphan receptors.

Chromatin immunoprecipitation (ChIP) assays in HepG2 cells indicate that PGC-1α is present near RORE on the proximal *Bmal1* promoter (Fig. 2d). In addition, PGC-1α recruits histone acetyltransferases, including p300 and GCN5, but not Clock, a recently described DNA-binding histone acetyltransferase[24], to this site. Histone hyperacetylation is associated with transcriptional activation. On the other hand, trimethylation of lysine 4 of histone 3 (H3K4me3) is a hallmark for actively transcribed genes, whereas H3K9 dimethylation (H3K9me2) is typically found in heterochromatin and silenced genes[25]. Remarkably, PGC-1α expression leads to a robust increase in histone H3 acetylation and H3K4me3 levels with a corresponding reduction in H3K9me2 levels on the proximal *Bmal1* promoter. These results indicate that PGC-1α activates *Bmal1* transcription by altering the local chromatin environment from a repressive to an active state.

The ability of PGC-1α to regulate the expression of clock genes indicates that this factor may impact on the circadian clock, thereby linking clock oscillators to energy metabolism. To test whether PGC-1α is required for clock function *in vivo*, we monitored locomotor activity and body temperature in wild-type and *PGC-1α* null mice

using telemetry. Wild-type mice exhibit a free-running period of approximately 23.7 h in constant darkness, whereas the free-running period is approximately 24.0 h in PGC-1α null mice (Fig. 3a). In addition, locomotor activity is evenly distributed throughout the subjective night in the null mice. These results indicate that PGC-1α regulates locomotor behaviour in mice and is required for maintaining normal circadian periods. Diurnal oscillation of body temperature is also distinct between these two genotypes (Fig. 3b). Because PGC-1α is an important regulator of energy metabolism, we next examined whether the daily oscillation of metabolic rate is altered in the null mice. Using indirect calorimetry, we measured oxygen consumption rate (VO_2) in wild-type and *PGC-1α* null mice over a period of three days[6]. Consistent with locomotor activity and body temperature rhythms, VO_2 is significantly higher in the dark phase (Fig. 3c). Compared with wild-type mice, *PGC-1α* null mice have elevated VO_2 in both light and dark phases, though the diurnal oscillation of VO_2 is severely blunted. These results indicate that PGC-1α is a critical component of the clock circuitry that may link the clock oscillator to metabolism.

Analysis of clock gene expression indicates that the rise of *Bmal1* mRNA levels is significantly impaired (37% lower at CT19), although the rhythmic expression of *Bmal1* is nearly intact in *PGC-1α* null livers (Fig. 3d). In contrast, peak *Bmal1* mRNA level is higher in PGC-1α-deficient skeletal muscle (Fig. 3e), indicating that PGC-1α deficiency differentially perturbs clock genes in different tissues. Diurnal rhythms of *Clock* and *Per1* mRNA are also disrupted in the liver and skeletal muscle of *PGC-1α* null mice, respectively. We next examine whether PGC-1α is required for the cyclic expression of various enzymes involved in energy metabolism. *PEPCK (PCK1)*, a rate-limiting enzyme of hepatic gluconeogenesis, shows a daily cycle that peaks at CT13 in the wild-type liver. PGC-1α seems to be dispensable for the rhythmic expression of this enzyme. Nevertheless, the mRNA levels of PEPCK are constitutively elevated at CT1 and CT13 (Fig. 3d), in a similar way to the previous studies[6]. Pyruvate dehydrogenase kinase 4 (PDK4) is a key regulator of the pyruvate dehydrogenase complex that controls the flux of pyruvate into the tricarboxylic acid cycle. The diurnal *PDK4* expression is severely impaired in PGC-1α-deficient liver and skeletal muscle, although the phase of PDK4 oscillation seems to be preserved. Similarly, the cyclic expression of fatty acid binding protein 3 (*Fabp3/H-FABP*) as well as other mitochondrial genes involved in oxidative phosphorylation (OXPHOS), such as aconitase (*Aco2*) and *Cox4a* (*Cox4i1*), is also disrupted in *PGC-1α* null skeletal muscle. Together, our results illustrate that PGC-1α is essential for oscillatory expression of clock and metabolic genes *in vivo*.

The defects in circadian rhythms of clock and metabolic gene expression in *PGC-1α* null mice could potentially be affected by altered central clock function. Previous work indicates that peripheral oscillators can be reset by restricted feeding, essentially uncoupling them from the master pacemaker in the suprachiasmatic nucleus[26]. To determine whether PGC-1α is required for phase resetting of peripheral clocks in response to feeding, we subjected wild-type and null mice to night feeding and then switched them to day feeding. Blood glucose concentrations show robust oscillation when the mice were fed exclusively at night (Fig. 4a). The phase of glucose oscillation is shifted by approximately 12 h after the feeding switch in both groups. However, the amplitude of blood glucose rhythms is significantly dampened in mice lacking PGC-1α. qPCR analyses revealed that the expression of *PGC-1α* and clock genes are reversed in response to restricted feeding in wild-type livers (Fig. 4b). In contrast, feeding-induced resetting of clock gene expression is significantly disrupted in *PGC-1α* null liver and skeletal muscle (Fig. 4b and Supplementary Fig. 2). The expression of *Bmal1* in the suprachiasmatic nucleus is also reduced in *PGC-1α* null mice (Supplementary Fig. 2); however, the feeding switch does not seem to alter *Bmal1* expression patterns. Interestingly, *PGC-1β* mRNA levels are elevated at some timepoints in these tissues.

To rule out the possibility that the perturbation of peripheral clocks in *PGC-1α* null mice may be secondary to a compromised central clock, we examined cell-autonomous clock function *in vivo*

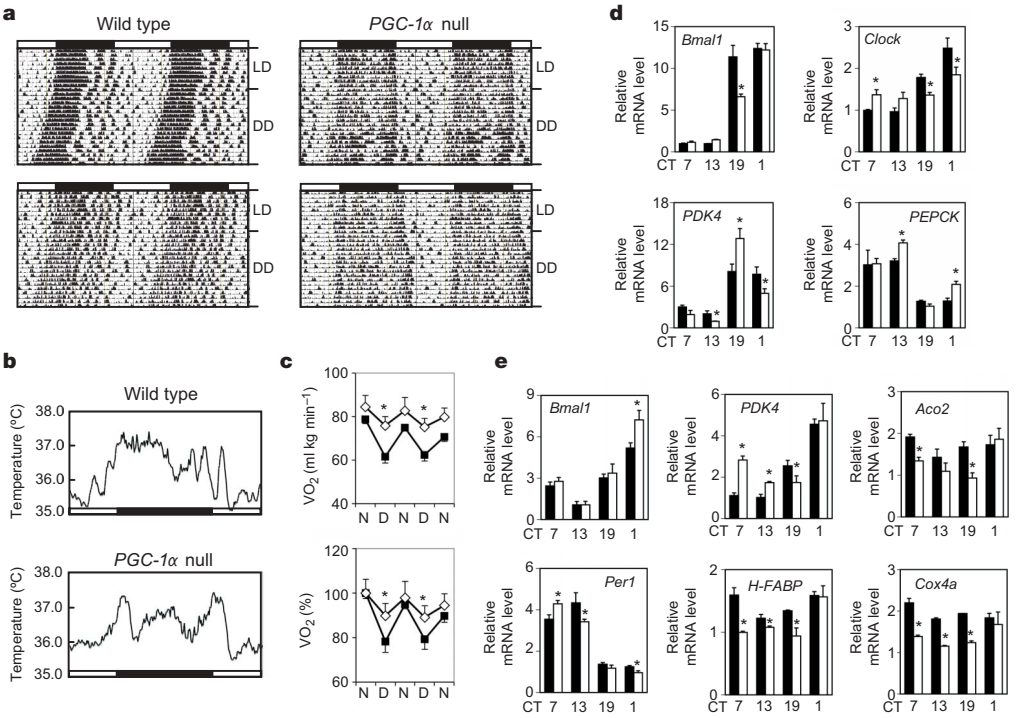


Figure 3 | Disruption of clock function in *PGC-1α* null mice. **a**, Locomotor activity in wild-type and *PGC-1α* null mice under normal light:dark cycles 12:12 h (LD) or constant darkness (DD). Data were plotted in duplicate columns in each panel. **b**, Representative traces of core body temperature in wild-type and *PGC-1α* null mice. **c**, Metabolic rate in wild-type (filled square, $n = 5$) and *PGC-1α* null (open diamond, $n = 6$) mice. Shown are average and normalized VO_2 during subjective night (N) and day (D). Error bars, s.e.m. **d**, **e**, qPCR analysis of gene expression in pooled liver (**d**) and skeletal muscle (**e**). Error bars, s.d. of the mean. $^*P < 0.05$, wild-type (filled box) versus null mice (open box).

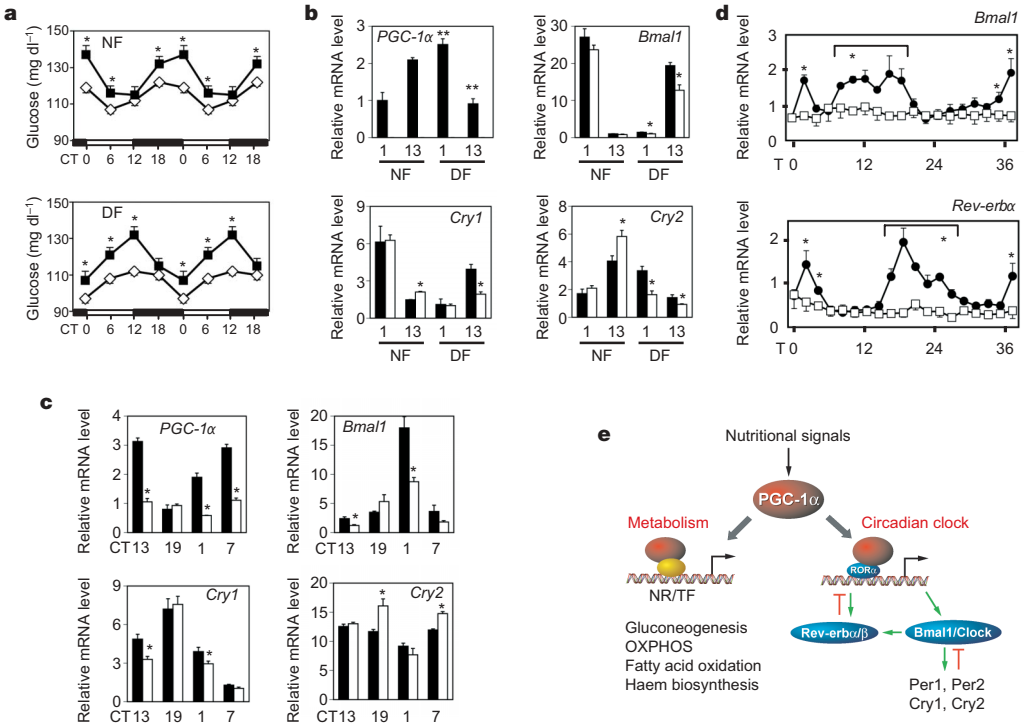


Figure 4 | Cell-autonomous role of PGC-1α in clock regulation. **a**, Plasma glucose concentrations in wild-type (filled square, $n = 16$) and *PGC-1α* null mice (open diamond, $n = 14$) subjected to night (NF) or day (DF) feeding. Data were plotted in duplicate columns. **b**, qPCR analysis of total liver RNA isolated from wild-type (filled box) and *PGC-1α* null (open box) mice. **c**, qPCR analyses of hepatic gene expression in mice transduced with control (filled box) or *PGC-1α* RNAi (open box) adenoviruses. Error bars in **a**–**c** indicate s.e.m. **a**, **b**, **c**, $^{*}P < 0.05$ wild-type versus null mice; $^{**}P < 0.01$ NF versus DF. **d**, Time course expression of clock genes in wild-type (filled circle) and *PGC-1α* null (open square) fibroblasts following serum shock. Error bars, s.d. $^{*}P < 0.05$ wild-type versus null cells. **e**, Model for coordinated regulation of circadian clock and energy metabolism through PGC-1α.

as well as in cultured cells. C57/Bl6J mice were transduced through the tail vein with adenoviruses expressing random or short hairpin RNA (shRNA) directed towards *PGC-1α*[27]. RNA interference (RNAi) knockdown of *PGC-1α* in the liver significantly disrupts rhythmic expression of *Bmal1*, *Cry1* and *Cry2* in this tissue, but not in the suprachiasmatic nucleus (Fig. 4c and Supplementary Fig. 2). Serum shock has been demonstrated to induce rhythmic clock gene expression in cultured fibroblasts[28]. We exposed immortalized brown preadipocytes to brief serum shock and analysed expression of clock genes over a period of 36 h. Serum shock leads to robust oscillation of *Bmal1* and *Rev-erbα* expression in wild-type cells (Fig. 4d). In contrast, rhythmic expression of these genes is essentially abolished in cells lacking PGC-1α. Consistent with these results, the ability of RORα to induce the expression of *Bmal1* and *Rev-erbα* is impaired in hepatocytes lacking PGC-1α (Supplementary Fig. 3), indicating that PGC-1α is required for the transcriptional activity of RORα in this context. We conclude from these studies that PGC-1α exerts its effects on the circadian clock largely in a tissue- and cell-autonomous manner.

We have identified PGC-1α, a major metabolic regulator, as a critical component of the mammalian clock. PGC-1α stimulates the expression of *Bmal1* through coactivating the ROR family of orphan nuclear receptors and is essential for normal circadian rhythms. Because PGC-1α expression is highly responsive to nutritional signals and potentially light, our findings support a mechanism through which energy metabolism and circadian clock can be directly coupled at the transcriptional level (Fig. 4e). *PGC-1α* null mice are resistant to diet-induced obesity and are more insulin-sensitive[6,17], whereas *Clock* mutant mice develop obesity. These differences are probably due to the fact that *PGC-1α* null mice are hyperactive and have a higher metabolic rate in the absence of increased food intake. Our findings raise an intriguing possibility that the expression and/or activity of PGC-1α itself may be regulated by components of the clock oscillator. Because feeding and locomotor activity are regulated by circadian clocks, it is possible that rhythmic PGC-1α expression is controlled by these physiological and behavioural rhythms. Alternatively, PGC-1α is associated with the SirT1 histone deacetylase complex[29] and may directly sense the metabolic state of the cell, in a similar way to the regulation of the clock homologue NPAS2 by redox status[30]. Disruption of circadian rhythms has been implicated in the pathogenesis of metabolic disorders. Our studies have uncovered a potential molecular target that could simultaneously modulate circadian clocks and energy metabolism.

METHODS SUMMARY

Transcriptional analyses. Adenoviral transduction, reporter gene assays, and ChIP experiments were performed as previously described [6,9]. For gene expression analyses, total RNA was isolated from tissues or cultured cells using Trizol reagents (Invitrogen), reversed transcribed and analysed by quantitative PCR using SYBR-Green.

Animal experiments. For activity and body temperature monitoring, wild-type and *PGC-1α* null mice were implanted with a G2 Minimiter II probe in the abdominal cavity (MiniMitter) and allowed to recover for two weeks prior to the initiation of monitoring using telemetric receivers. Voluntary activity and body temperature were recorded under light:dark cycles 12:12 h for two weeks before switching to constant darkness for three more weeks. For restricted feeding, male wild-type and *PGC-1α* null mice were fed exclusively at night for ten days. Blood glucose levels were measured at the end of night feeding as well as four days following the switch to day feeding. Tissues were harvested at CT1 and 13 for both feeding groups for qPCR analysis of gene expression. For liver-specific PGC-1α knockdown, wild-type C57/Bl6J male mice were administered control or *PGC-1α* RNAi adenoviruses via the tail vein. Tissues were harvested from transduced animals five days after adenoviral transduction.

Serum shock. Immortalized fibroblasts from wild-type and *PGC-1α* null mice were established and maintained in DMEM supplemented with 10% FBS. For serum shock, media of confluent cultures was replaced with DMEM plus 50% horse serum ($t = 0$) and, after 1 hour, the cells were washed once with PBS and

incubated with serum-free DMEM. Total RNA was extracted at the indicated time points and processed for qPCR analysis. More details are provided in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 January; accepted 21 March 2007.
Published online 2 May 2007.

1. Rutter, J., Reick, M. & McKnight, S. L. Metabolism and the control of circadian rhythms. *Annu. Rev. Biochem.* **71,** 307–331 (2002).
2. Schibler, U. & Naef, F. Cellular oscillators: rhythmic gene expression and metabolism. *Curr. Opin. Cell Biol.* **17,** 223–229 (2005).
3. Kelly, D. P. & Scarpulla, R. C. Transcriptional regulatory circuits controlling mitochondrial biogenesis and function. *Genes Dev.* **18,** 357–368 (2004).
4. Leone, T. C. *et al.* PGC-1α deficiency causes multi-system energy metabolic derangements: muscle dysfunction, abnormal weight control and hepatic steatosis. *PLoS Biol.* **3,** e101 (2005).
5. Lin, J., Handschin, C. & Spiegelman, B. M. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab.* **1,** 361–370 (2005).
6. Lin, J. *et al.* Defects in adaptive energy metabolism with CNS-linked hyperactivity in PGC-1α null mice. *Cell* **119,** 121–135 (2004).
7. Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92,** 829–839 (1998).
8. Wu, Z. *et al.* Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* **98,** 115–124 (1999).
9. Yoon, J. C. *et al.* Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413,** 131–138 (2001).
10. Lowrey, P. L. & Takahashi, J. S. Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu. Rev. Genomics Hum. Genet.* **5,** 407–441 (2004).
11. Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109,** 307–320 (2002).
12. Storch, K. F. *et al.* Extensive and divergent circadian gene expression in liver and heart. *Nature* **417,** 78–83 (2002).
13. Ueda, H. R. *et al.* A transcription factor response element for gene expression during circadian night. *Nature* **418,** 534–539 (2002).
14. Wijnen, H. & Young, M. W. Interplay of circadian clocks and metabolic rhythms. *Annu. Rev. Genet.* **40,** 409–448 (2006).
15. Karlsson, B., Knutsson, A. & Lindahl, B. Is there an association between shift work and having a metabolic syndrome? Results from a population based study of 27,485 people. *Occup. Environ. Med.* **58,** 747–752 (2001).
16. Rudic, R. D. *et al.* BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS Biol.* **2,** e377 (2004).
17. Turek, F. W. *et al.* Obesity and metabolic syndrome in circadian Clock mutant mice. *Science* **308,** 1043–1045 (2005).
18. Schibler, U. & Sassone-Corsi, P. A web of circadian pacemakers. *Cell* **111,** 919–922 (2002).
19. Reppert, S. M. & Weaver, D. R. Coordination of circadian timing in mammals. *Nature* **418,** 935–941 (2002).
20. Preitner, N. *et al.* The orphan nuclear receptor REV-ERBα controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110,** 251–260 (2002).
21. Sato, T. K. *et al.* A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron* **43,** 527–537 (2004).
22. Yin, L., Wang, J., Klein, P. S. & Lazar, M. A. Nuclear receptor Rev-erbα is a critical lithium-sensitive component of the circadian clock. *Science* **311,** 1002–1005 (2006).
23. Lau, P., Nixon, S. J., Parton, R. G. & Muscat, G. E. RORα regulates the expression of genes involved in lipid homeostasis in skeletal muscle cells: caveolin-3 and CPT-1 are direct targets of ROR. *J. Biol. Chem.* **279,** 36828–36840 (2004).
24. Doi, M., Hirayama, J. & Sassone-Corsi, P. Circadian regulator CLOCK is a histone acetyltransferase. *Cell* **125,** 497–508 (2006).
25. Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nature Rev. Mol. Cell Biol.* **6,** 838–849 (2005).
26. Damiola, F. *et al.* Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev.* **14,** 2950–2961 (2000).
27. Koo, S. H. *et al.* PGC-1 promotes insulin resistance in liver through PPAR-α-dependent induction of TRB-3. *Nature Med.* **10,** 530–534 (2004).
28. Balsalobre, A., Damiola, F. & Schibler, U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93,** 929–937 (1998).
29. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1α and SIRT1. *Nature* **434,** 113–118 (2005).
30. Rutter, J., Reick, M., Wu, L. C. & McKnight, S. L. Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. *Science* **293,** 510–514 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Hogenesch and T. Sato for *Bmal1-luc*, *RORα* and *Rev-erb* plasmids; S. Reppert for the E-box luciferase constructs; J. Takahashi for the *Bmal1* and *Clock* expression plasmids; and M. Montminy for *PGC-1α* RNAi adenovirus. We also thank A. Saltiel and B. Spiegelman for comments on the manuscript and P.-H. Wu for discussions. This work is supported by NIDDK (J.D.L.) and the University of Michigan BSSP program (J.D.L.).

Author Contributions C.L., S.L., J.B. and J.D.L. designed the research. C.L., S.L., T.L., and J.D.L. performed the experiments. C.L., S.L., T.L., J.B. and J.D.L. analysed the data. J.D.L. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.D.L. (jdlin@umich.edu).

METHODS

Adenoviral transduction. The maintenance and adenoviral transduction of primary hepatocytes and C2C12 myotubes was performed, as previously described[6]. Total RNA was isolated using Trizol reagents (Invitrogen) 48 h following transduction, reverse transcribed, and analysed by qPCR using SYBR Green. A complete list of PCR primers is shown in Supplementary Table 2. Primers for ribosomal protein 36B4 were included for normalization. Microarray analysis was performed on total RNA isolated from hepatocytes transduced with *GFP* or *PGC-1α* adenoviruses using U74av2 chips (Affymetrix).

Animals. All animal procedures were approved by the University Committee on Use and Care of Animals. For analysis of PGC-1α expression in tissues, male C57/Bl6J mice of 12 weeks of age were kept under light:dark (LD) 12:12 h and subsequently subjected to constant darkness for 36 h. Tissues from five mice were dissected every 4 h for a total of 24 h thereafter. Total RNA was isolated and analysed by qPCR using gene-specific primers. For analysis of gene expression in wild-type and *PGC-1α* null mice, three to five mice of each genotype were euthanized every six hours. Tissues were immediately frozen and subsequently processed for qPCR analysis and immunoblotting analyses.

For activity and body temperature monitoring, 3-month old wild-type and *PGC-1α* null mice were implanted with a G2 Minimiter II probe in abdominal cavity (MiniMitter). Animals were allowed to recover for two weeks prior to the initiation of monitoring using telemetric receivers. Voluntary activity and body temperature were recorded under LD 12:12 h for two weeks before switching to constant darkness for three more weeks. Data were analysed using software supplied by the manufacturer. For restricted feeding, male wild-type and *PGC-1α* null mice were fed exclusively at night for ten days. Blood glucose levels were measured at the end of night feeding as well as four days following the switch to day feeding. Tissues were harvested at CT1 and 13 for both feeding groups for qPCR analysis of gene expression. For liver-specific PGC-1α knockdown, a total of 32 C57/Bl6J male mice were administered control or *PGC-1α* RNAi adenoviruses (0.1 absorbance units per mouse), as previously described. Five days following tail vein injection, tissues were harvested from transduced animals at CT1, 7, 13 and 19 (four mice per treatment).

Reporter gene assays. Reporter gene assays were performed in BOSC 293 cells, as previously described. In a typical experiment, 25 ng of reporter plasmids were mixed with 10 ng of expression constructs for transcription factors in the presence or absence of PGC-1α expression construct. Equal amounts of DNA were used for all transfection combinations by adding appropriate vector DNA. Relative luciferase activities were determined 48 h following transfection. All transfection experiments were performed in triplicates.

Serum shock. Immortalized fibroblasts from wild-type and *PGC-1α* null mice were established and maintained in DMEM supplemented with 10% FBS. For serum shock, media of confluent cultures was replaced with DMEM plus 50% horse serum ($t = 0$), and after 1 h, the cells were washed once with PBS and incubated with serum-free DMEM. Total RNA was extracted at the indicated time points and processed for qPCR analysis using 36B4 as a normalization control.

ChIP assay. Chromatin immunoprecipitation was performed essentially as described by the Upstate Biotechnology (www.upstate.com). Briefly, HepG2 hepatoma cells were transduced with *GFP* or *Flag–PGC-1α* adenoviruses for 48 h. Chromatin lysates were prepared, pre-cleared with Protein-A agarose beads, and immunoprecipitated with antibodies against Flag (Sigma), K9-dimethylated and K4-trimethylated histone H3 (Abcam), acetylated histone H3 (Upstate Biotechnology), p300 and Clock (Santa Cruz Biotech), GCN5 (BioLegend), or normal mouse IgG (Sigma) in the presence of BSA and salmon sperm DNA. Beads were extensively washed before reverse cross-linking. DNA was purified using a PCR purification kit (Qiagen) and subsequently analysed by PCR using primers flanking the proximal RORE on the human *Bmal1* promoter, as detailed in Supplementary Table 2.

LETTERS

Disulphide-isomerase-enabled shedding of tumour-associated NKG2D ligands

Brett K. Kaiser[1]*, Daesong Yim[1]*, I-Ting Chow[1]*, Segundo Gonzalez[1]†, Zhenpeng Dai[1], Henning H. Mann[1], Roland K. Strong[1], Veronika Groh[1] & Thomas Spies[1]

Tumour-associated ligands of the activating NKG2D (natural killer group 2, member D; also called KLRK1) receptor—which are induced by genotoxic or cellular stress—trigger activation of natural killer cells and co-stimulation of effector T cells, and may thus promote resistance to cancer[1–6]. However, many progressing tumours in humans counter this anti-tumour activity by shedding the soluble major histocompatibility complex class-I-related ligand MICA, which induces internalization and degradation of NKG2D and stimulates population expansions of normally rare $NKG2D^+CD4^+$ T cells with negative regulatory functions[7–9]. Here we show that on the surface of tumour cells, MICA associates with endoplasmic reticulum protein 5 (ERp5; also called PDIA6 or P5), which, similar to protein disulphide isomerase, usually assists in the folding of nascent proteins inside cells[10]. Pharmacological inhibition of thioreductase activity and *ERp5* gene silencing revealed that cell-surface ERp5 function is required for MICA shedding. ERp5 and membrane-anchored MICA form transitory mixed disulphide complexes from which soluble MICA is released after proteolytic cleavage near the cell membrane. Reduction of the seemingly inaccessible disulphide bond in the membrane-proximal α3 domain of MICA must involve a large conformational change that enables proteolytic cleavage. These results uncover a molecular mechanism whereby domain-specific deconstruction regulates MICA protein shedding, thereby promoting tumour immune evasion, and identify surface ERp5 as a strategic target for therapeutic intervention.

NKG2D-mediated tumour rejection can be effective at early stages of tumour growth[1–3,5]. However, sustained surface expression and shedding of soluble MICA (sMICA) by late-stage human tumours negatively imprint on the local and systemic immune response, thus promoting tumour immune evasion[7–9]. The significance of this relationship is highlighted by beneficial effects of neutralizing anti-MICA antibodies that were induced as a result of immunotherapy in a clinical trial[11]. These findings are supported by a mouse model study that confirmed systemically impaired natural killer cell and CD8 T-cell functions, accompanied by increased tumour susceptibility, as a result of NKG2D downmodulation by locally sustained NKG2D ligand expression[12]. However, mice lack sequences that are orthologous to human *MICA* and the closely related *MICB*, and shedding of naturally expressed NKG2D ligands has not been observed in mice[6].

Early studies of the interactions between NKG2D and its ligands used randomly oligomerized recombinant MICA or ULBP family ligands produced as immunoglobulin fusion proteins, all of which bound exclusively to NKG2D-expressing lymphocytes[13,14]. On testing MICA and ULBP2 tetramers, we confirmed that these high-avidity reagents stained the NKL natural killer cell line and that binding was entirely accounted for by NKG2D (Fig. 1a). However, during the course of screening ~40 cell lines by flow cytometry, we observed that MICA, but not ULBP2, tetramers stained cell types lacking NKG2D. The highest fluorescence intensities were recorded with U266 myeloma cells and all of the ten epithelial tumour lines tested, and correlated with relatively large amounts of cell-surface MICA. Only monocytic U937 cells were identified as negative for tetramer binding (Fig. 1a). The MICA tetramers were prepared using glycosylated protein secreted by transfected 293T cells. However, tetramer binding was not reduced after cleavage of cell-surface polypeptide-linked carbohydrates but was inhibited in the presence of unglycosylated recombinant MICA (Fig. 1a). Thus, these results revealed an interaction involving MICA—but not NKG2D ligands in general—and an unidentified surface protein.

Candidate MICA-binding proteins were purified from U266 and negative control U937 outer cell membranes using MICA-coupled sepharose beads. SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and silver staining revealed two sets of protein bands that were detected with U266 but not U937 cells (Fig. 1b). By mass spectrometry, two protein bands in the 76–78-kilodalton (kDa) molecular mass range corresponded to heat shock 70 kDa protein 5 (glucose-regulated protein, 78 kDa) (HSPA5, also known as BiP). A major protein band of 50 kDa was identified as ERp5 (ref. 10), and two additional proteins of about 47 and 48 kDa shared similarities with thioredoxin family members. Because all of these proteins are typically intracellular, we scrutinized their outer cell membrane localization. Using the same purification protocol and surface-biotinylated cells, immunoblots probed with streptavidin-horse radish peroxidase (HRP) or polyclonal antibodies demonstrated the presence of HSPA5, and more prominently ERp5, on the surface of U266 but not U937 cells, which was confirmed by staining for ERp5 (Fig. 1c, d). ERp5 is related to protein disulphide isomerase. Both proteins contain two thioredoxin-like domains, each with a pair of active site cysteines in CXXC motifs, and mediate the intracellular formation of nascent polypeptide disulphide bonds; however, they have also been implicated in extracellular disulphide exchange[10,15–17].

In exploring a functional relationship between MICA and ERp5, we were guided by the epithelial tumour-associated expression that is characteristic of MICA but not the ULBP family of NKG2D ligands[6,9]. This idea was encouraged when freshly isolated tumour cells displayed similar patterns of tetramer and anti-ERp5 and anti-MICA antibody binding, and matched serum samples were positive for sMICA (Fig. 2a). We thus tested for a role of ERp5 in MICA shedding by exposing U266 cells and tumour lines HeLa, A375 melanoma, and HCT116 and LoVo colon carcinoma to DTNB (5,5-dithiobis-(2-nitrobenzoic acid)) or PAO (phenylarsine oxide), which impair

[1]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. †Present address: Biologia Funcional, Universidad de Oviedo, IUOPA, Oviedo, Asturias 33006, Spain.
*These authors contributed equally to this work.

protein disulphide isomerase function by forming disulphide and coordination bonds, respectively, with thiol groups in its catalytic sites[18]. Both inhibitors reduced the production of sMICA at titred non-toxic concentrations without affecting surface expression of MICA (Fig. 2b and data not shown). Treatment with PAO also diminished tetramer binding, suggesting that MICA interacts directly with an ERp5 catalytic site (Fig. 2c). However, these inhibitors are relatively nonspecific and may have pleiotropic effects. Therefore, and to preclude an involvement of thiol isomerases other than ERp5, we expressed short interfering (si)RNA constructs targeting two regions of *ERp5* messenger RNA in A375 cells. As measured by real-time reverse transcription PCR (RT–PCR), *ERp5* mRNA was reduced by ~70–80% as a result of the siRNA targeting (Fig. 3a). As a consequence, ERp5 surface expression, MICA tetramer binding and sMICA shedding decreased, although the amount of MICA surface protein was not noticeably changed (Fig. 3b, c). Thus, the cumulative evidence indicated that surface ERp5 function is required for MICA shedding.

The functional association between ERp5 and MICA was biochemically analysed. Initial failure to co-immunoprecipitate these proteins from lysates of surface-biotinylated HeLa cells implied that ERp5 and MICA maintain no stable complexes after solubilization. However, ERp5 co-immunoprecipitated with MICA when HeLa cells were treated with trichloroacetic acid (TCA), which traps mixed disulphide polypeptides and quenches thiol interchange (Fig. 4a, lane 1)[19]. Sulphydryl groups in subsequent cell lysates were alkylated and immunocomplexes deglycosylated with *N*-glycanase. This procedure was modified by using HeLa cells grown in the presence of denatured and reduced RNase (dRNase), which served as excess substrate, shifting ERp5 equilibrium towards the reduced state[20] and thereby favouring disulphide exchange with MICA. Analysis by SDS–PAGE and immunoblotting showed that increasing the concentration of dRNase resulted in larger amounts of ERp5 co-immunoprecipitating with MICA (Fig. 4a, lanes 1–4). Concomitantly, the MICA polypeptide of 38 kDa (shortened in HeLa cells owing to homozygous cytoplasmic tail deletion[21]) disappeared, and proteins with molecular masses of 31 and 34 kDa emerged. The 34-kDa protein corresponded to truncated sMICA, as determined by secondary precipitation from dissociated immunocomplexes and comparison to sMICA isolated from HeLa cell culture media (Fig. 4a, lanes 3, 4, 9 and 12). The 31-kD protein may represent another substrate or co-factor that was recruited into ERp5–MICA complexes. Similar data were obtained using anti-ERp5 for immunoprecipitations (Fig. 4a, lanes 5 and 6). None of those biochemical changes was observed when cells were grown in the presence of native RNase (Fig. 4b). Thus, these results

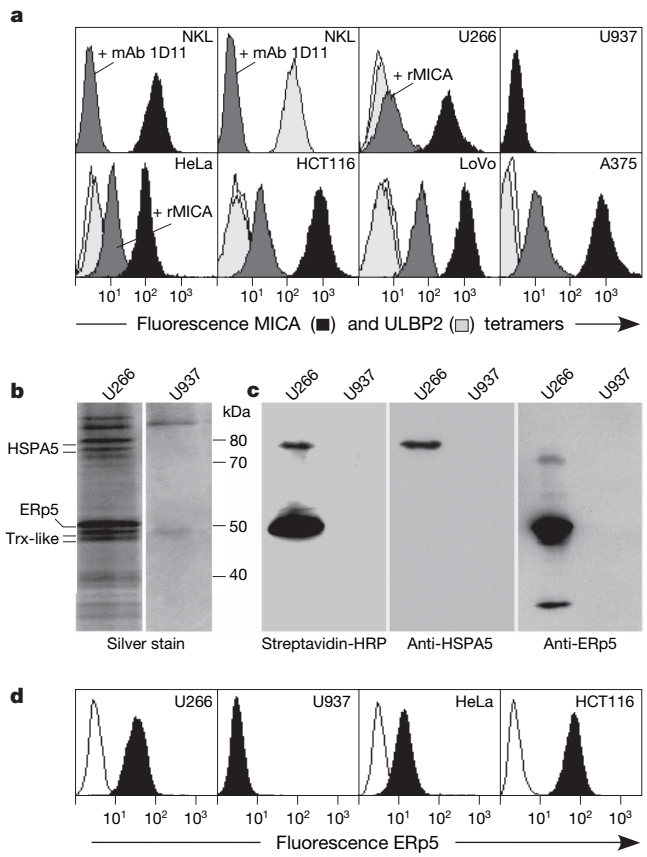


Figure 1 | Surface interactions of MICA with ERp5 and HSPA5. a, Flow cytometry confirms MICA (black shading) and ULBP2 (light-grey shading) tetramer binding to NKL cells and inhibition by anti-NKG2D monoclonal antibody 1D11 (dark-grey shading). MICA but not ULBP2 tetramers bind to NKG2D-negative U266, HeLa, HCT116, LoVo and A375 tumour cells, and binding is inhibited by bacterial recombinant MICA (dark-grey shading). U937 cells are negative for MICA tetramer binding. Open profiles are control IgG stainings. **b**, Silver staining of U266 and U937 outer cell membrane proteins enriched for binding to MICA beads. Trx, thioredoxin. **c**, Probing of MICA bead-purified proteins from surface biotinylated cells with streptavidin-HRP or specific antisera. Two additional bands in the anti-ERp5 lane are cross-reactive. **d**, Anti-ERp5 stainings (black shading). Open profiles are IgG controls.

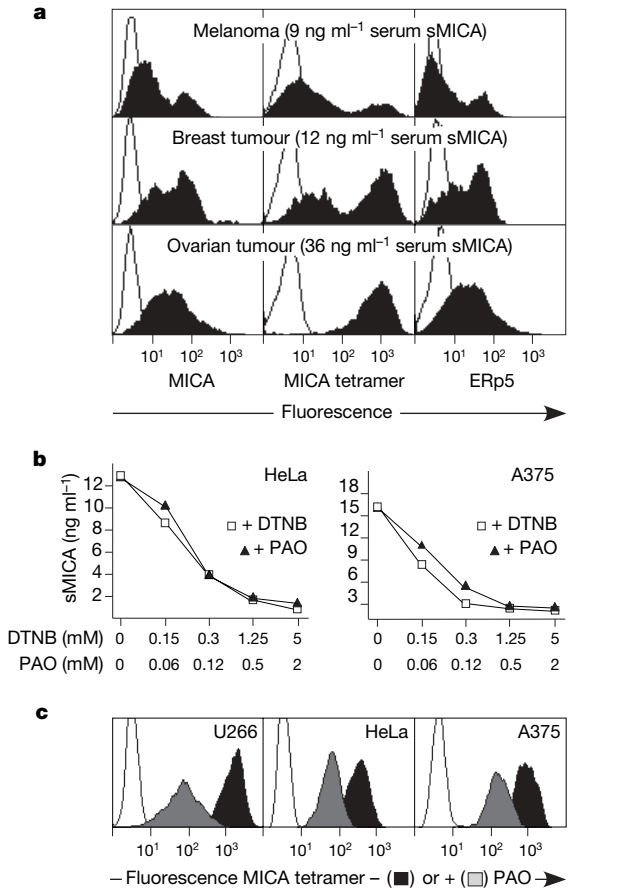


Figure 2 | Tumour-associated ERp5 surface expression and pharmacological inhibition of sMICA shedding. a, Freshly isolated melanoma, breast and ovarian tumour cells are positive for MICA, tetramer binding and surface ERp5 (black shading; open profiles are IgG control stainings for anti-MICA and anti-ERp5, and phycoerythrin-streptavidin controls for MICA tetramer). Matched patient peripheral blood serum samples contain the indicated amounts of sMICA. Data are representative of five matched sample pairs. **b**, DTNB and PAO reduce shedding of sMICA by HeLa and A375 cells in a dose-dependent manner, as determined by ELISA. Similar results were obtained with HCT116 and LoVo cells. **c**, PAO interferes with MICA tetramer binding. Open profiles are phycoerythrin-streptavidin controls.

demonstrated dynamic interactions between ERp5 and MICA that were closely tied to the production of sMICA, which was corroborated by large increases of sMICA in dRNase-treated HeLa and A375 cell cultures (Fig. 4c).

To demonstrate ERp5-mediated MICA disulphide bond reduction and explore substrate and domain specificities, bacterially produced recombinant proteins were mixed and incubated in the absence of reducing agents and thus under oxidizing conditions. Non-reducing gel electrophoresis and comparison to β-mercaptoethanol (β-ME)-treated samples showed gradual reduction of MICA (Supplementary Fig. 1). This was remarkable as ERp5 affected an intact, properly folded, substrate protein isoenergetically and in the absence of any other factor in solution. A similar result was obtained with the closely related MICB (data not shown). In contrast, ERp5 did not affect unrelated proteins with relatively accessible intra-chain (siderocalin) or intra-chain and inter-chain (KLRD1–NKG2A) disulphide bonds (Supplementary Fig. 1). No synergistic effect was observed when MICA was exposed to ERp5 together with HSPA5. Of the two ERp5 thioredoxin-like domains, which were expressed as two separate polypeptides (amino acid residues 1–118 and 135–421; Fig. 5a), only the amino-terminal domain displayed functional activity (Fig. 5b and data not shown). As with protein disulphide isomerase, ERp5 uses a catalytic mechanism whereby one active site cysteine invades the target disulphide, transiently forming a disulphide-linked heterodimer that is resolved by disulphide exchange with the second active site cysteine[10,15]. Of two ERp5(1–118) mutant fragments, the C36S point mutation showed no activity on MICA substrate whereas C39S formed a trapped disulphide-linked intermediate, thus confirming the role of C36 as the invading and C39 as the resolving cysteine in this reaction (Fig. 5c). By size-exclusion chromatography, intact ERp5 was a trimer in solution whereas the two individual domains behaved as monomers (data not shown). Thus, ERp5 multimerization was not required for MICA reduction.

Similar to conventional major histocompatibility complex class I molecules, MICA contains three intra-chain disulphide bonds located between amino acid residues 36 and 41, 96 and 164, and 202 and 259 in the α1, α2 and the C-type immunoglobulin-like α3 domain, respectively[22]. To identify the target disulphide, the α1α2 platform and α3 membrane-proximal domains were expressed and tested separately. ERp5 displayed no activity with the α1α2 domain (Fig. 5d). Because we were unable to electrophoretically resolve reduced and non-reduced forms of the relatively small α3 domain, we used the ERp5(1–118) polypeptide fragment with the C39S mutation for analysis. Gel electrophoresis revealed a large protein band shift corresponding to a mixed disulphide heterodimer (Fig. 5e). Thus, the disulphide bond targeted by ERp5 was in the MICA α3 domain.

Proteolytic cleavage of MICA is thought to be mediated by metalloproteinases[23]. However, with HeLa and A375 cells we observed no reduction in sMICA shedding by metalloproteinase inhibitors, suggesting that diverse proteases may have the ability to cleave MICA. To determine the MICA cleavage site, we purified sMICA from cultures of transfectant C1R-MICA cells, which express modest amounts of ERp5 but proliferate vigorously in serum-free media. Carboxy-terminal sequencing by mass spectroscopic analysis of tryptic peptide fragments revealed ragged MICA C termini defined by several neighbouring amino acid residues at or near the transmembrane boundary.

Our results suggest that MICA cleavage occurs in complex with ERp5 before mixed disulphide resolution, which in all likelihood results in immediate 'snap-back' oxidization of the MICA disulphide bond. ERp5 escape from intracellular retention is probably independent of MICA, as intracellular interactions have not been observed[24]. Precedent for biological functions of surface thiol isomerases includes

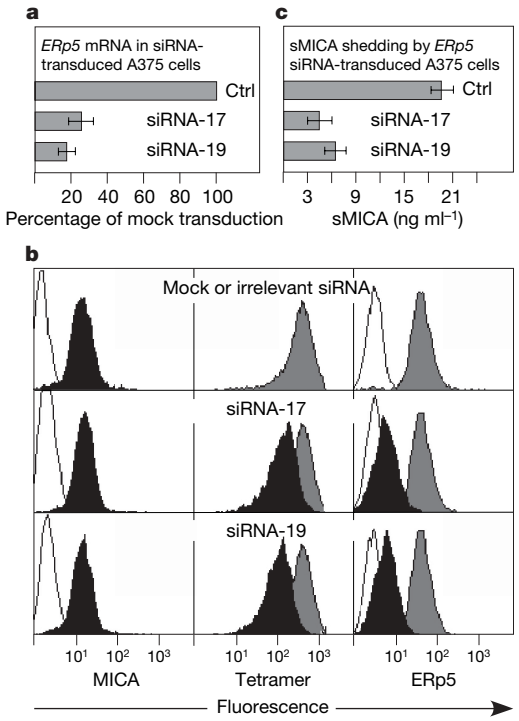


Figure 3 | ERp5 is required for sMICA shedding. **a**, Expression of siRNA constructs 17 or 19 in A375 cells results in ~70–80% reduction of *ERp5* mRNA, as measured by real-time RT–PCR. **b**, Knockdown of *ERp5* mRNA decreases MICA tetramer binding and ERp5 surface expression (black shading in centre and right columns). MICA expression (black shading in left column) is unchanged; open profiles represent IgG control stainings. **c**, Knockdown of *ERp5* mRNA diminishes sMICA shedding as determined by ELISA. Control bars in **a** and **c** and grey-shaded control profiles in **b** represent mock-transduced cells or cells expressing irrelevant siRNA. Error bars in **a** and **c** represent deviations among three experiments.

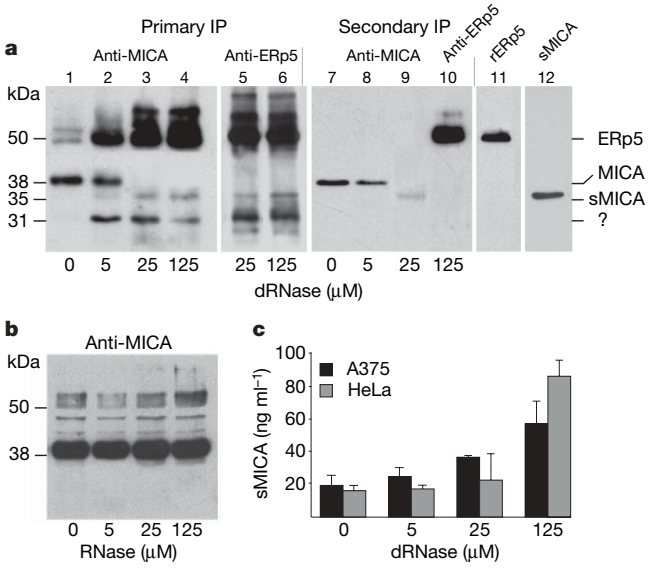


Figure 4 | ERp5–MICA disulphide exchange enables MICA cleavage. **a**, Treatment of surface-biotinylated HeLa cells with TCA before lysis, MICA immunoprecipitation, SDS–PAGE and membrane transfer reveals MICA–ERp5 complexes (lane 1). Protein identities are confirmed by secondary precipitations (lanes 7, 8, 10), primary precipitations of ERp5 (lanes 5, 6) and by recombinant ERp5 (lane 11). After cell culture in the presence of dRNase, co-immunoprecipitated ERp5 increases (lanes 2–4), full-length MICA disappears and sMICA emerges (lanes 3, 4). sMICA identity is confirmed by secondary precipitation (lane 9) and comparison to sMICA from cell culture media (lane 12). **b**, Control experiment with cells grown in the presence of native RNase. **c**, dRNase promotes sMICA shedding. Error bars represent deviations among three experiments.

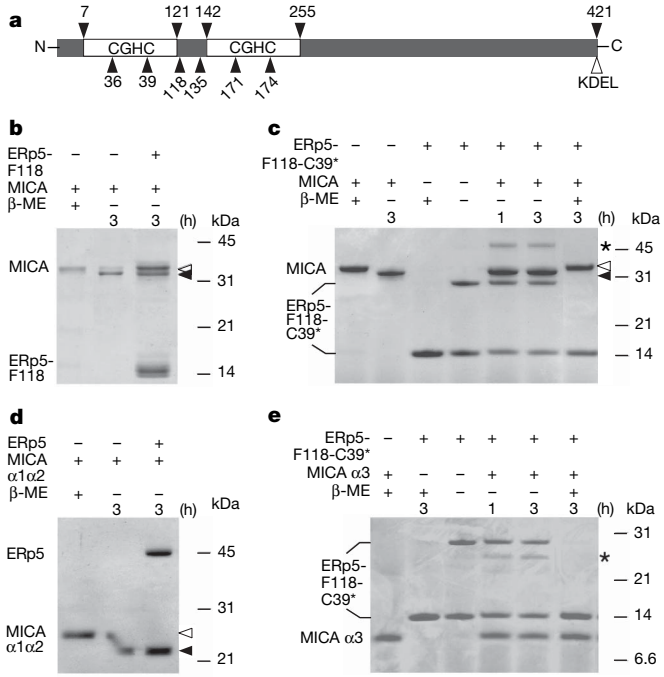


Figure 5 | ERp5 exhibits specificity for the MICA α3 domain. a, ERp5 organization with CGHC motifs within thioredoxin domains (open boxes). The top row of numbers identifies amino acid positions at domain boundaries; the bottom row of numbers identifies cysteine positions and truncation sites of expressed ERp5 fragments. **b**, MICA is partially reduced by ERp5(1–118) (ERp5-F118). **c**, MICA and the C39S mutant (C39*) of ERp5(1–118) form mixed disulphide heterodimers (asterisk) that are resolved by β-ME. The C39S mutant of ERp5(1–118) partially forms homodimers (see also **e**). **d**, ERp5 has no effect on the MICA α1α2 domain. The partial reduction in lane 2 is owing to bleeding of β-ME from lane 1. **e**, The C39S mutant of ERp5(1–118) reduces the MICA α3 disulphide bond as indicated by unresolved heterodimers (asterisk). Filled and open arrowheads in **b**–**d** mark the positions of non-reduced and reduced forms, respectively, of MICA substrates.

alteration of integrin affinity states, CD4 homodimer formation by inter-chain disulphide exchange, which enables HIV-1 T-cell infection, and switching of cell-surface tissue factor functional states between activation of coagulation and G-protein-coupled signalling[16,17,25–27]. The function of ERp5 demonstrated here enables tumour immune evasion and may influence autoimmune diseases through sMICA-mediated T-cell modulation[6].

METHODS SUMMARY

Tetramers, antibodies, protein identification and ELISA for sMICA. Tetramers were prepared from recombinant proteins, which were expressed in transfected 293T cells and purified using Invitrogen methodology, by BirA enzymatic biotinylation and conjugation with phycoerythrin-streptavidin. Anti-NKG2D (monoclonal antibody 1D11) and anti-MICA (monoclonal antibody 2C10) antibodies have been described[13,24]. Rabbit anti-ERp5 and anti-HSPA5 were from Affinity BioReagents and Stressgen, respectively. MICA binding proteins were purified from U266 cells by dounce-homogenization, dextran-PEG partitioning, Triton X-114 phase separation, and affinity chromatography using MICA-conjugated sepharose beads, followed by analysis of separated protein bands by mass spectrometry. For immunoblotting, MICA-binding proteins were prepared after cell-surface biotinylation with EZ-Link Sulfo-NHS-LC-Biotin (Pierce). The enzyme-linked immunosorbent assay (ELISA) for sMICA has been described[7].

siRNA expression and real-time RT–PCR. Retroviral transduction using pBABE-GFP constructs and Phoenix amphotropic packaging cells were used for *ERp5* siRNA expression. As with real-time RT–PCR[9], oligonucleotide sequences and further details are given in the Methods section.

Immunoprecipitations and MICA cleavage. Denatured and reduced RNase A was prepared as described[20]. HeLa cells were exposed to dRNase for 16 h, washed and surface biotinylated, incubated in 10% (w/v) TCA in phosphate-buffered saline (PBS) for 30 min on ice, washed, and lysed with immediate pH neutralization in standard NP-40 lysis buffer containing protease inhibitors and *N*-ethylmaleimide. Immunoprecipitated protein complexes were treated with *N*-glycanase and subjected to SDS–PAGE and membrane transfer. C-terminal truncation analysis of sMICA was performed by mass spectrometry at the Harvard University Microchemistry Facility.

ERp5 activity assays. All proteins, domains and the C36S and C39S mutants (made by Stratagene Quick Change methodology) were expressed in bacteria and purified as described[28]. ERp5 and substrate proteins were incubated at room temperature in PIPES buffer and resolved in Tris-glycine or Bis-Tris NuPAGE (Invitrogen) gels.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 March; accepted 22 March 2007.
Published online 9 May 2007.

1. Cerwenka, A., Baron, J. L. & Lanier, L. L. Ectopic expression of retinoic acid early inducible-1 gene (RAE-1) permits natural killer cell-mediated rejection of a MHC class I-bearing tumor *in vivo*. *Proc. Natl Acad. Sci. USA* **98,** 11521–11526 (2001).
2. Diefenbach, A., Jensen, E. R., Jamieson, A. M. & Raulet, D. H. Rae1 and H60 ligands of the NKG2D receptor stimulate tumor immunity. *Nature* **413,** 165–171 (2001).
3. Girardi, M. *et al.* Regulation of cutaneous malignancy by γδ T cells. *Science* **294,** 605–609 (2001).
4. Gasser, S., Orsulic, S., Brown, E. J. & Raulet, D. H. The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature* **436,** 1186–1190 (2005).
5. Smyth, M. J. *et al.* NKG2D function protects the host from tumor initiation. *J. Exp. Med.* **202,** 583–588 (2005).
6. Gonzalez, S., Groh, V. & Spies, T. Immunobiology of human NKG2D and its ligands. *Curr. Top. Microbiol. Immunol.* **298,** 121–138 (2006).
7. Groh, V., Wu, J., Yee, C. & Spies, T. Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature* **419,** 734–738 (2002).
8. Doubrovina, E. S. *et al.* Evasion from NK cell immunity by MHC class I chain-related molecules expressing colon carcinoma. *J. Immunol.* **171,** 6891–6899 (2003).
9. Groh, V., Smythe, K., Dai, Z. & Spies, T. Fas ligand-mediated paracrine T cell regulation by the receptor NKG2D in tumor immunity. *Nature Immunol.* **7,** 755–762 (2006).
10. Ellgaard, L. & Ruddock, L. W. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep.* **6,** 28–32 (2005).
11. Jinushi, M., Hodi, F. S. & Dranoff, G. Therapy-induced antibodies to MHC class I chain-related protein A antagonize immune suppression and stimulate antitumor cytotoxicity. *Proc. Natl Acad. Sci. USA* **103,** 9190–9195 (2006).
12. Oppenheim, D. E. *et al.* Sustained localized expression of ligand for the activating NKG2D receptor impairs natural cytotoxicity *in vivo* and reduces tumor immunosurveillance. *Nature Immunol.* **6,** 928–937 (2005).
13. Bauer, S. *et al.* Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285,** 727–729 (1999).
14. Cosman, D. *et al.* ULBPs, novel MHC class I-related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor. *Immunity* **14,** 123–133 (2001).
15. Kikuchi, M., Doi, E., Tsujimoto, I., Horibe, T. & Tsujimoto, Y. Functional analysis of human P5, a protein disulfide isomerase homologue. *J. Biochem.* **132,** 451–455 (2002).
16. Turano, C., Coppari, S., Altieri, F. & Ferraro, A. P. Proteins of the PDI family: unpredicted non-ER localizations and functions. *J. Cell. Physiol.* **193,** 154–163 (2002).
17. Jordan, P. A. & Gibbins, J. M. Extracellular disulfide exchange and the regulation of cellular function. *Antioxid. Redox Signal.* **8,** 312–324 (2006).
18. Gallina, A. *et al.* Inhibitors of protein-disulfide isomerase prevent cleavage of disulfide bonds in receptor-bound glycoprotein 120 and prevent HIV-1 entry. *J. Biol. Chem.* **277,** 50579–50588 (2002).
19. Frand, A. R. & Kaiser, C. A. Ero1p oxidizes protein disulfide isomerase in a pathway for disulfide bond formation in the endoplasmic reticulum. *Mol. Cell* **4,** 469–477 (1999).
20. Essex, D. W., Chen, K. & Swiatkowska, M. Localization of protein disulfide isomerase to the external surface of the platelet plasma membrane. *Blood* **86,** 2168–2173 (1995).
21. Groh, V. *et al.* Broad tumor-associated expression and recognition by tumor-derived γδ T cells of MICA and MICB. *Proc. Natl Acad. Sci. USA* **96,** 6879–6884 (1999).
22. Li, P. *et al.* Crystal structure of the MHC class I homolog MIC-A, a γδ T cell ligand. *Immunity* **10,** 577–584 (1999).
23. Salih, H. R., Rammensee, H.-G. & Steinle, A. Down-regulation of MICA on human tumors by proteolytic shedding. *J. Immunol.* **169,** 4098–4102 (2002).
24. Groh, V. *et al.* Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc. Natl Acad. Sci. USA* **93,** 12445–12450 (1996).

25. Matthias, L. J. *et al.* Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nature Immunol.* **3,** 727–732 (2002).
26. Ahamed, J. *et al.* Disulfide isomerization switches tissue factor from coagulation to cell signaling. *Proc. Natl Acad. Sci. USA* **103,** 13932–13937 (2006).
27. Maekawa, A., Schmidt, B. & Fazekas de St.. Groth, B. Sanejouand, Y.-H. & Hogg, P. J. Evidence for a domain-swapped CD4 dimer as the coreceptor for binding to class II MHC. *J. Immunol.* **176,** 6873–6878 (2006).
28. Li, P. *et al.* Complex structure of the actiavting immunoreceptor NKG2D and its MHC class I-like ligand MICA. *Nature Immunol.* **2,** 443–451 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Carter for protein purification advice; M. Welcker for help with siRNA expression; W. Lane and P. Gafkan for mass spectrometry analyses; K. Smythe for technical assistance; and S. Riddell for comments on the manuscript. This work was supported by the Cancer Research Institute (B.K.K.), the Spanish Fondo de Investigaciones Sanitarias (S.G.), the Deutsche Forschungsgemeinschaft (H.H.M.), the Edson Fund, the Avon Foundation Breast Cancer Immunotherapy Research Initiative, and by grants from the NIH.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.S. (tspies@fhcrc.org).

METHODS

Tumour samples and cell lines, antibodies, tetramers, pharmacological inhibitors, and ELISA for sMICA. The source of tumour cell suspensions has been reported previously[7]. Cell lines were from the American Type Culture Collection. Anti-NKG2D (monoclonal antibody 1D11) and anti-MICA (monoclonal antibody 2C10) antibodies have been described[13,24]. Rabbit polyclonal anti-ERp5 and anti-HSPA5 antibodies were from Affinity BioReagents and Stressgen, respectively. Recombinant MICA*001 (residues 1–276) and ULBP2 (residues 1–202) were produced in transfected 293T cells and purified from culture supernatant using Invitrogen methodology. Tetramers were prepared by BirA enzymatic biotinylation and conjugation with phycoerythrin-streptavidin. Cells were stained with saturating tetramer concentrations for 1 h at 4 °C and examined by flow cytometry. Non-glycosylated MICA was expressed in bacteria and purified as described[22]. Cells were exposed to DTNB or PAO (Sigma) for 24 h at the indicated concentrations. For inhibition of tetramer binding, cells were grown for 24 h in the presence of 0.5 μM PAO. Metalloproteinase inhibitors GM 6001 and MMP inhibitor III were from Calbiochem. The ELISA for sMICA has been described[7].

Identification of MICA-binding surface proteins. U266 and U937 cells (each 5×10^9) were dounce-homogenized in 10 mM Tris-HCl (pH 7.6), 0.5 mM $MgCl_2$, 1 mM PMSF, 1 μg ml^{-1} leupeptin, and 1 μg ml^{-1} pepstatin. Membrane fractions were isolated from cleared supernatants by dextran-PEG partitioning, washed (8% sucrose, 5 mM Tris-HCl (pH 7.4)), and dissociated in lysis buffer (50 mM Tris-Cl (pH 7.4), 1% Triton X-114, 150 mM NaCl, 5 mM EDTA, 5 mM iodoacetamide, protease inhibitors). Cleared supernatants were warmed to 37 °C and proteins partitioned during Triton X-114 phase separation. Proteins in aqueous fractions were affinity purified using MICA conjugated to cyanogen-bromide-activated sepharose beads, visualized by SDS–PAGE and silver staining, and analysed by MALDI-TOF at the FHCRC Mass Spectrometry Facility. For immunoblotting, MICA-binding proteins were prepared after cell surface biotinylation with EZ-Link Sulfo-NHS-LC-Biotin (Pierce).

siRNA expression and real-time RT–PCR. Oligonucleotide pairs for siRNA-17 and siRNA-19 targeting *ERp5* (disulphide isomerase-related protein P5; GenBank accession number D49489) mRNA at positions 316–338 and 556–567 were GATCTTGTTGTCAAAGTTGGTGCAGTTGTCTTCTTCTCAACTG-CACCAACTTTGACAACATTTTTG and AATTCAAAAATGTTGTCAAAG-TTGGTGCAGTTGAGAAGAAGACAACTGCACCAACTTTGACAACAA, and GATCTTGATAGTTCAAGTAAGAAGGATGTCTTCTTCTCATCCTTCTTA-CTTGAACTATCATTTTTG and AATTCAAAAATGATAGTTCAAGTAAGA-AGGATGAGAAGAAGACATCCTTCTTACTTGAACTATAA, respectively (all 5′–3′; internal hairpin sequence, 3′-end termination signal, and *Bgl*II and *Eco*RI overhangs are underlined). An irrelevant oligonucleotide pair with no homology to any human gene was GATCTTATGTCAAGTTGTATAGTTA-TTCAAGAGATAACTATACAACTTGACATATTTTTG and AATTCAAAAA-TATGTCAAGTTGTATAGTTATCTCTTGAATAACTATACAACTTGACATAA. Annealed primers were ligated into retroviral vector pBABE-GFP and constructs were sequenced. Virus was produced in Phoenix amphotropic packaging cells and culture supernatant used for infection of A375 cells, which were sorted for GFP expression. Real-time RT–PCR was performed as described[9], using primer sets TGCGGCACGCTGCAGGGCT and TTGACAGTGACCACACCATGGAGCATA for *ERp5* complementary DNA, and GGAACGGAAAGGACCTCAGGATG and CTGGGAGCTCCTGGTGCTGTTG for *MICA* cDNA, and SYBR Green reagents (Molecular Probes).

Preparation of dRNase and capturing of ERp5–MICA complexes. RNase A was denatured and reduced in 0.1 M Tris-OH (pH 8.6), 6 M guanidine hydrochloride, and 0.15 M dithiothreitol for 24 h at room temperature, and desalted using D-Salt Dextran columns (Pierce) equilibrated with PBS[20]. 2×10^6 semi-confluent HeLa cells were exposed to the indicated concentrations of dRNase or native RNase for 16 h, washed, and surface biotinylated with EZ-Link Sulfo-NHS-LC-Biotin (Pierce). Labelled cells were incubated in 0.5 ml 10% (w/v) TCA in PBS for 30 min on ice, washed sequentially in 10% and 5% TCA in PBS, and lysed in 50 mM Tris (pH 7.4), 1% Surfact-Amps NP-40 (Pierce), 150 mM NaCl, 5 mM EDTA, 40 mM *N*-ethylmaleimide (Sigma), 1 mM PMSF, leupeptin (1 μg ml^{-1}), and pepstatin (1 μg ml^{-1}). Lysate pH was adjusted to 7.0 with 1 M Tris-OH (pH 9.5). Protein complexes were precipitated with monoclonal antibody 2C10 (anti-MICA) or ERp5 polyclonal antibody, treated with *N*-glycanase, and processed for SDS–PAGE. For sequential precipitation, monoclonal antibody 2C10 immunocomplexes were dissociated in 150 mM Tris (pH 7.4), 0.5% SDS and 10 mM dithiothreitol, diluted tenfold with lysis buffer containing 25 mM iodoacetamide, incubated for 1 h at room temperature for dithiothreitol neutralization and sulphydryl alkylation, and re-precipitated with anti-MICA monoclonal antibody BAMO-1 (Axxora) or anti-ERp5. For determination of sMICA C-terminal cleavage, supernatant from C1R-MICA transfectants grown in Opti-MEM (Gibco) was concentrated using Amicon Ultra-15 centrifugal filters (Millipore). Immunoprecipitated sMICA was treated with *N*-glycanase, isolated by SDS–PAGE, and subjected to peptide fragmentation analysis by MALDI mass spectrometry at the Harvard University Microchemistry Facility.

ERp5 activity assays. Ectodomain-only MICA, siderocalin and KLRD1–NKG2A were expressed in bacteria and purified as described[28]. We similarly produced ERp5 (residues 1–421 of the mature protein), the ERp5 fragments 1–118 and 135–421, the C36S and C39S mutants (made by Stratagene Quick Change methodology) of ERp5(1–118), and the isolated MICA α1α2 platform (residues 1–180) and α3 domains (residues 187–274)[22]. All *ERp5* sequences were fused to N-terminal hexahistidine tracts and included a C-terminal stop codon to prevent expression of the adjacent hexahistidine in pET22(b). Recombinant proteins were purified by metal affinity (BD Talon, Clontech) and size-exclusion chromatography (Superdex 200, Pharmacia). For testing of functional activity, ERp5 or derivative proteins (2 μg) were incubated at room temperature with MICA substrates or control proteins (1.5 μg) in PNEA (25 mM PIPES (pH 7), 150 mM NaCl, 1 mM EDTA, and 0.02% sodium azide) in a total volume of 5 μl per time point sample, mixed with 2x SDS–PAGE sample buffer (5 μl) with or without β-ME, and resolved in 15% Tris-glycine or 12% Bis-Tris NuPAGE (Invitrogen) gels.

LETTERS

Redox-mediated substrate recognition by Sdp1 defines a new group of tyrosine phosphatases

G. C. Fox[1]*†, M. Shafiq[1]*, D. C. Briggs[1], P. P. Knowles[1], M. Collister[4], M. J. Didmon[4], V. Makrantoni[5], R. J. Dickinson[4], S. Hanrahan[2], N. Totty[2], M. J. R. Stark[5], S. M. Keyse[4] & N. Q. McDonald[1,3]

Reactive oxygen species trigger cellular responses by activation of stress-responsive mitogen-activated protein kinase (MAPK) signalling pathways[1,2]. Reversal of MAPK activation requires the transcriptional induction of specialized cysteine-based phosphatases that mediate MAPK dephosphorylation[3]. Paradoxically, oxidative stresses generally inactivate cysteine-based phosphatases by thiol modification and thus could lead to sustained or uncontrolled MAPK activation[4,5]. Here we describe how the stress-inducible MAPK phosphatase, Sdp1, presents an unusual solution to this apparent paradox by acquiring enhanced catalytic activity under oxidative conditions. Structural and biochemical evidence reveals that Sdp1 employs an intramolecular disulphide bridge and an invariant histidine side chain to selectively recognize a tyrosine-phosphorylated MAPK substrate. Optimal activity critically requires the disulphide bridge, and thus, to the best of our knowledge, Sdp1 is the first example of a cysteine-dependent phosphatase that couples oxidative stress with substrate recognition. We show that Sdp1, and its paralogue Msg5, have similar properties and belong to a new group of phosphatases unique to yeast and fungal taxa.

MAPK phosphatases (MKPs) have a Cys-$(X)_5$-Arg catalytic motif and a conserved general acid Asp, both characteristic of cysteine-based phosphatases (CBPs)[6,7]. Most MKPs are dual-specific for phosphothreonine and phosphotyrosine, although phosphotyrosine-specific MKPs have also been identified[3,8,9]. Each can target and dephosphorylate phospho-residues within the MAPK activation loop thereby inhibiting MAPK activity[3]. MKPs are often transcriptionally induced following stimulation of their target MAPK as part of a negative feedback loop to attenuate MAPK signalling[3]. When the activating stimulus for a MAPK involves reactive oxygen species, this poses a conundrum as, like all CBPs, MKPs are potentially sensitive to oxidative inactivation[5,10]. Protective mechanisms have been described for CBPs that prevent irreversible oxidative damage to the catalytic cysteine, either through a sulphenyl-amide intermediate (PTP1B) or by formation of a reversible intramolecular disulphide bond with the catalytic cysteine (PTEN, CDC25, KAP, MKP-3)[5,11,12]. In each case the phosphatase is inactivated, albeit reversibly. In yeast, adaptation to oxidative stress is a necessary requisite for survival and multiple MAPKs are known to respond to oxidative stimuli, including Slt2/Mpk1 (refs 1, 13, 14). Moreover, strains lacking Slt2 are hypersensitive to oxidants[14–17]. We predicted that MKPs specific for Slt2, which are required to operate under oxidative conditions, may employ mechanisms to deal with this environmental stress. Slt2 forms part of a signalling cascade required to maintain cell wall integrity and can also respond to other stresses such as heat shock[18]. Regulation of Slt2 involves at least four distinct MKPs able to dephosphorylate Slt2 directly[13]. However, only the stress-dependent phosphatase Sdp1, induced by peroxide stimulus, is specific for Slt2 (refs 19, 20). We therefore investigated how Sdp1 is able to dephosphorylate a MAPK substrate under oxidative conditions.

In the presence of the reducing agent dithiothreitol (1 mM DTT) Sdp1 exhibited a relatively poor K_m (Michaelis constant) of 14.5 mM towards the chromogenic phosphatase substrate *para*-nitrophenyl phosphate (pNPP) (Fig. 1a). However, in the absence of reducing agents, Sdp1 has a K_m value of 0.6 mM, indicating a 24-fold enhanced affinity for pNPP substrate and therefore a much higher catalytic efficiency (Fig. 1a). The k_{cat} for pNPP remained approximately the same in the presence or absence of reducing agent. In marked contrast, the VHR phosphatase, typical of other CBPs, required reducing agents to maintain its reactive thiol in a reduced form (Fig. 1a).

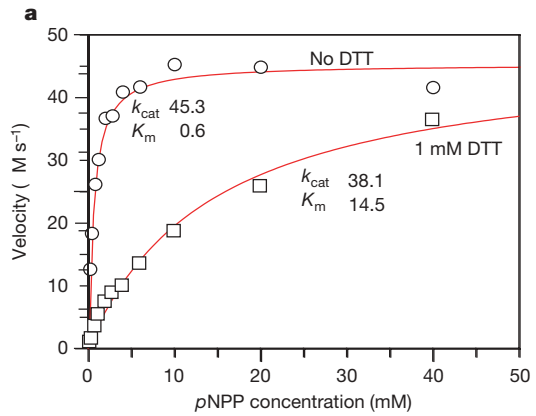
Sdp1 has been shown to dephosphorylate both activated Slt2 and the closely related mammalian ERK2 MAPK[19,20]. Western blots using anti-phospho-specific antibodies showed that Sdp1 exclusively targets the phosphotyrosine residue of activated ERK2 known to be required for catalytic activity (Fig. 1b). Remarkably, the absence of reducing agent has a dramatic effect on the kinetics of phosphotyrosine dephosphorylation by Sdp1 (Fig. 1b). All ERK2 phosphotyrosine is hydrolysed within 10 min without DTT, whereas total hydrolysis is not achieved even after 2 h of incubation in the presence of 1 mM DTT (Fig. 1b). Both the dual-specificity MKP-3 and tyrosine-specific VHR phosphatases were able to dephosphorylate activated ERK2 fully under these conditions[21,22] (Fig. 1c). Thus, Sdp1 is the first reported cysteine-based phosphatase that is essentially inactivated by reducing agents.

Dual-specificity MKPs that use an amino-terminal non-catalytic domain in MAPK substrate recognition have been described[3], and therefore we assayed a truncated Sdp1 spanning approximately the core phosphatase domain (56–197; Fig. 1c). This protein exhibited no activity towards phospho-ERK2 with or without DTT, suggesting that critical residue(s) were missing (Fig. 1c). To identify the region of Sdp1 required for ERK2 dephosphorylation and optimal pNPP activity we performed a deletion analysis from the amino terminus and found that residues 46–55 were essential (Fig. 1d). Sdp1 has only three cysteine residues (Fig. 1c) and one of these, Cys 47, is within this region. Quantification of the number of free sulphydryl groups in Sdp1 using Ellman's reagent (DTNB) suggested only a single free sulphydryl group was available, which we presumed corresponded to the Cys 140 active site residue (Fig. 2a). A C140S active site mutant of Sdp1 exhibited no reactivity with DTNB indicating there were no available –SH groups, which is consistent with Cys 47 and Cys 142

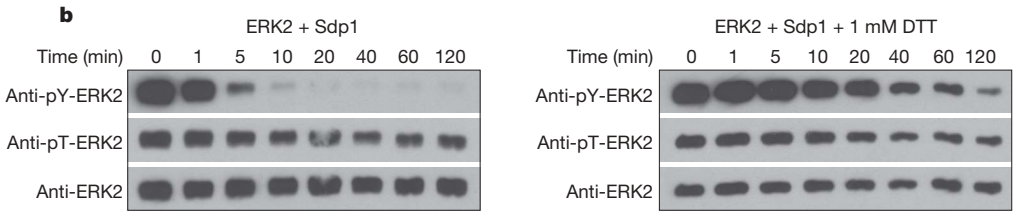
[1]Structural Biology Laboratory and [2]Protein Analysis Laboratory, The London Research Institute, Cancer Research UK, 44 Lincoln's Inn Fields, London WC2A 3PX, UK. [3]School of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK. [4]Cancer Research UK, Stress Response Laboratory, Biomedical Research Centre, Ninewells Hospital, Dundee DD1 9SY, UK. [5]Division of Gene Regulation and Expression, College of Life Sciences, MSI/WTB Complex, University of Dundee, Dundee DD1 5EH, UK. †Present address: ESRF, 6 Rue Jules Horowitz, 38043 Grenoble, Cedex 9, France.
* These authors contributed equally to this work.

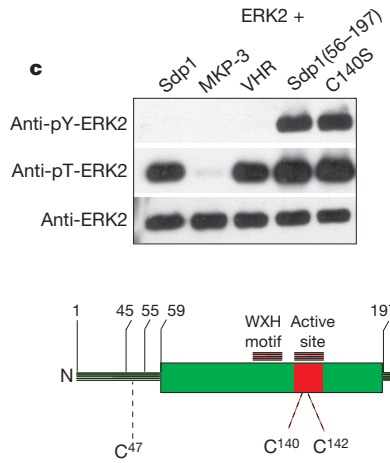
forming a disulphide bridge (Fig. 2a). To obtain direct evidence for an intramolecular Cys 47–Cys 142 disulphide bridge, we analysed a trypsin-digest of the Sdp1(C140S) mutant by matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry (MS) (Fig. 2b and Supplementary material). A disulphide-linked peptide at mass/charge ratio (*m/z*) 2,475 was identified and further fragmented by MS/MS CID (collision-induced dissociation). This generated two sets of triplet signals corresponding to symmetric (*m/z* 1,226 and 1,251) and asymmetric cleavage of the disulphide bridge linking the peptides EFVPCSSLDVR (residues 43–53) and ILIHSQCGLSR (residues 136–146) (Fig. 2b and Supplementary Table 1). Conversely, a reduced and alkylated Sdp1(C140S) mutant showed no *m/z* 2,475 ion after trypsin digestion but showed the equivalent cysteine peptide ions modified by alkylation (Supplementary Table 1). As expected from our biochemical analysis C47S, C142S as well as the C47S/C142S double Sdp1 mutant were all catalytically inactive towards phospho-ERK2 and exhibited only minimal basal activity towards pNPP (Fig. 2c and d). Moreover, these low-activity Sdp1 cysteine mutants were now insensitive to the presence of reducing agents (Fig. 2d).

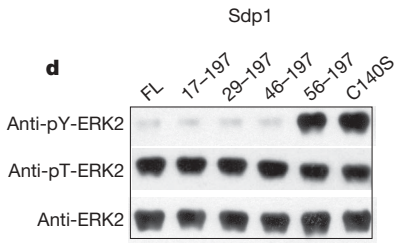
Intramolecular disulphide bond formation can result in increased migration of proteins analysed by non-reducing SDS–polyacrylamide



Protein	Kinetic constants (sec^{-1}) and (mM)	Catalytic efficiency
Sdp1(FL)	k_{cat} 49.2 ± 0.5 K_m 0.7 ± 0.03	70.3
Sdp1(FL) + DTT	k_{cat} 34.8 ± 0.5 K_m 12.5 ± 0.6	2.8
Sdp1	k_{cat} 45.3 ± 0.8 K_m 0.6 ± 0.05	75.5
Sdp1 + DTT	k_{cat} 38.1 ± 2.3 K_m 14.5 ± 1.8	2.6
VHR	k_{cat} 52.2 ± 0.7 K_m 32.0 ± 1.0	1.6
VHR + DTT	k_{cat} 117.0 ± 2.4 K_m 28.9 ± 1.4	4.0





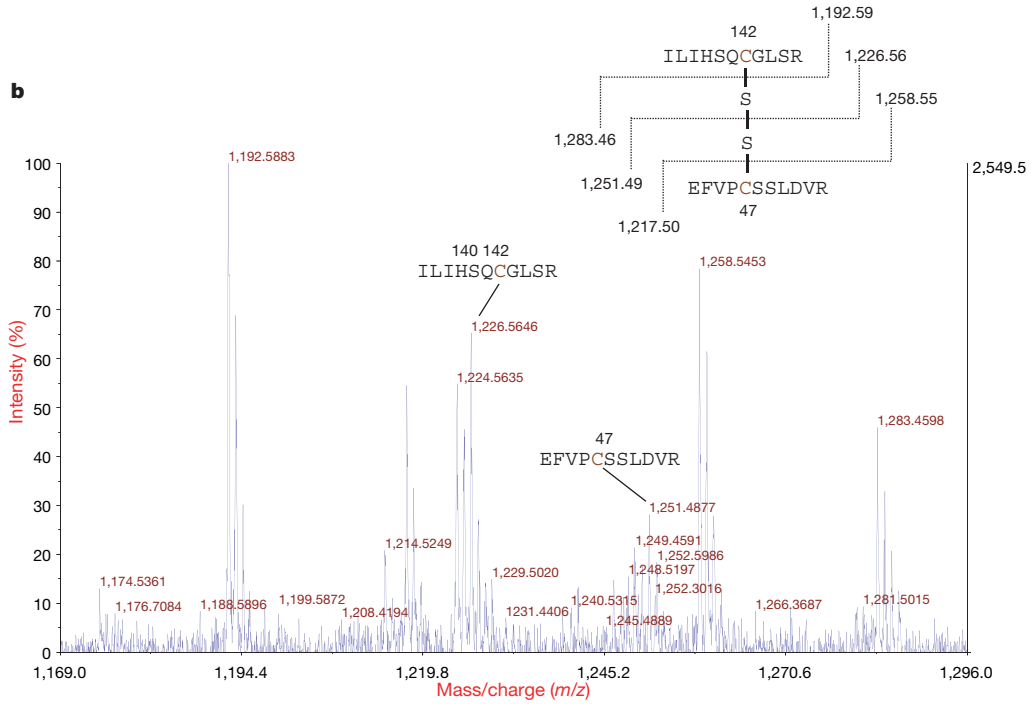


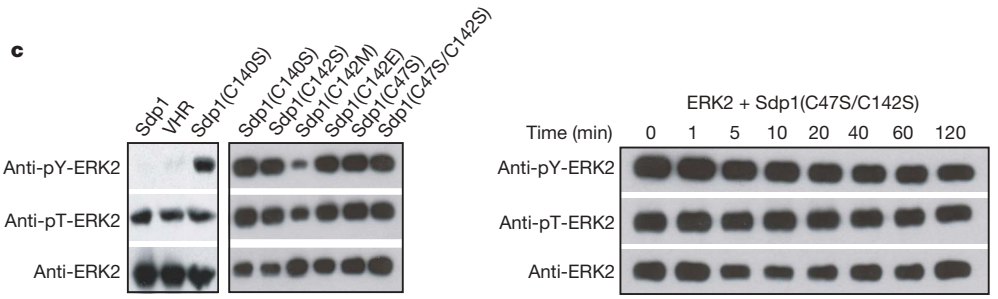
Protein	Kinetic constants (sec^{-1}) and (mM)	Catalytic efficiency
Sdp1	k_{cat} 44.6 ± 0.8 K_m 0.7 ± 0.13	63.7
Sdp1(29–197)	k_{cat} 44.1 ± 1.3 K_m 0.9 ± 0.1	49.0
Sdp1(46–197)	k_{cat} 31.9 ± 0.8 K_m 0.6 ± 0.07	53.2
Sdp1(56–197)	k_{cat} 11.5 ± 0.7 K_m 13.8 ± 1.8	0.8
C140S	Not detectable	None

Figure 1 | Sdp1 activity is sensitive to reducing agents. a, Effect of DTT on the Sdp1 kinetic constants for pNPP hydrolysis. Solid lines indicate fit to the Michaelis–Menten equation. Tabulated rate constants and s.e.m. for Sdp1 and VHR are also shown. Catalytic efficiency is defined as k_{cat}/K_m. Sdp1 is used to indicate a truncated form of Sdp1 (residues 17 to 197) used for structural and biochemical work as well as for the Sdp1 point mutants. Full-length Sdp1, indicated by Sdp1(FL), gave equivalent kinetic constants for both wild-type and mutant proteins. **b**, Time course of recombinant activated (diphosphorylated) ERK2 dephosphorylation by Sdp1 in the presence (right panel) or absence (left panel) of reducing agent, detected using phosphothreonine (pT) or phosphotyrosine (pY) specific antibodies by western blot analysis. **c**, Upper panel shows activated ERK2 dephosphorylation after a two-hour incubation with Sdp1 (no DTT) or the dual-specific DUSP6/MKP-3 and the tyrosine-specific VHR (both with 1 mM DTT). A truncated Sdp1 (residues 56–197) and a catalytically dead Sdp1(C140S) mutant were also tested without DTT. Lower panel shows a schematic for Sdp1, indicating the position of cysteine residues, and demarcates the catalytic CBP domain in green. **d**, Upper panel, activity of a series of N-terminal Sdp1 deletion mutants towards phospho-ERK2 by western blot, as previously described. Lower panel, same series of mutants and their activity towards pNPP.

a

Sample	Theoretical -SH	Measured -SH content
Sdp1	3	0.57 ± 0.04
Reduced Sdp1	3	3.11 ± 0.21
Sdp1 (C140S)	2	0.00
Reduced Sdp1 (C140S)	2	2.21 ± 0.10
BSA (17 disulphides)	35	0.57 ± 0.01
Glutathione	1	1.00 ± 0.02





d

Protein	Kinetic constants (sec^{-1}) and (mM)		Catalytic efficiency
Sdp1	k_{cat}	45.7 ± 0.8	57.1
	K_m	0.8 ± 0.06	
Sdp1 + DTT	k_{cat}	38.1 ± 2.3	2.6
	K_m	14.5 ± 1.8	
C47S/ C142S	k_{cat}	25.0 ± 0.7	2.1
	K_m	12.0 ± 1.0	
C47S/C142S + DTT	k_{cat}	29.0 ± 2.4	2.1
	K_m	14.0 ± 1.4	

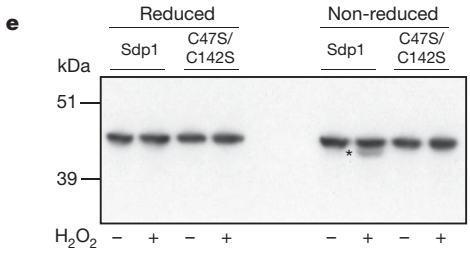


Figure 2 | Sdp1 activity requires an intramolecular Cys 47–Cys 142 disulphide bridge. **a**, Table of the free sulphydryl (-SH) content of wild type and various cysteine mutants of Sdp1 as measured by Ellman's reagent titration using glutathione and BSA as controls. **b**, MS/MS fragmentation spectrum of the Cys 47–Cys 142 disulphide-linked precursor ion at *m/z* 2,475. Inset, interpretation of the fragmentation. **c**, Activated ERK2 dephosphorylation by the Sdp1 cysteine point mutants (left panel) performed as described in the legend of Fig. 1c and a time course for the C47S/C142S double mutant (right panel). **d**, Kinetic rate constants for Sdp1 cysteine point mutants measured by pNPP hydrolysis. **e**, Analysis of the Sdp1 redox state *in vivo* by anti-Myc immunoblot. Sdp1 null yeast expressing either wild-type Myc–Sdp1 or Myc–Sdp1 C47S/C142S were subjected to heat shock at 39 °C for 45 min to induce Sdp1 expression and then either left untreated or exposed to H_2O_2 (0.4 mM) for 15 min. Protein extracts were then analysed by SDS–PAGE under non-reducing or reducing conditions, as indicated. The disulphide-containing Sdp1 protein with increased mobility is indicated by an asterisk.

gel electrophoresis (PAGE)[23]. To determine if the formation of an intramolecular disulphide bond within Sdp1 might be of physiological relevance, we have directly analysed the *in vivo* redox state of Sdp1 by SDS–PAGE and western blotting (Fig. 2e). As previously observed for the transcription factor Yap1, which forms an intramolecular disulphide bond when yeast cells are exposed to H_2O_2 (ref. 23), a proportion of the Sdp1 from H_2O_2-treated cells migrates more quickly than the protein from untreated cells under non-reducing conditions (lanes 5 and 6). However, this mobility shift was not detected after electrophoresis under reducing conditions (lanes 1 and 2). This shift is due to formation of an intramolecular disulphide, because the migration of Sdp1(C47S/C142S) was completely unaffected by exposure of yeast to H_2O_2 when analysed under non-reducing conditions (lanes 7 and 8). We also detect a distinct electrophoretic mobility shift when recombinant wild-type Sdp1 is analysed in the absence of DTT, which also depends on the integrity of Cys 47 and Cys 142, but not Cys 140 (see Supplementary Material and Supplementary Fig. 3).

To confirm our biochemical observations we determined the crystal structures of Sdp1 alone, bound to sulphate and phosphotyrosine as substrate mimetics (Fig. 3a). Superposition of the three molecules shows the equivalent general acid loop of CBPs in three discrete conformations corresponding to an open form with a 3_{10} helix, a partially closed form in the presence of sulphate in which the 3_{10} helix 'melts' and a fully closed state bound to phosphotyrosine (Fig. 3a, left panel). These changes reveal how the His 111 sidechain in the CBP acid loop moves into position only when a phosphotyrosine substrate occupies the active site. The structural changes are analogous to those observed for PTP1B and proposed as part of the activation mechanism for MKP-3, but involve a histidine rather than an aspartic acid[24,25].

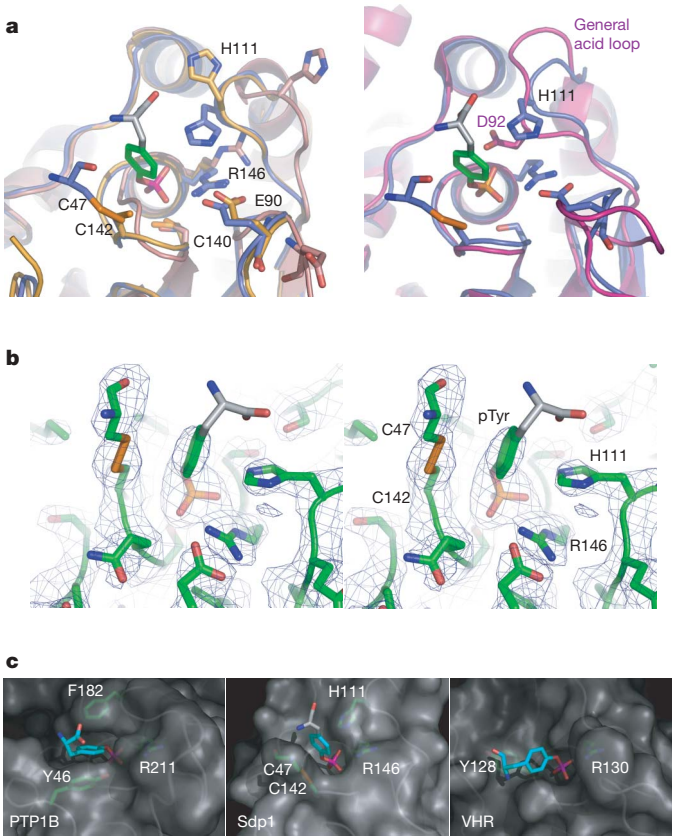


Figure 3 | Phosphotyrosine recognition by His 111 and the Cys 47–Cys 142 disulphide bridge. a, Left panel shows substrate-induced conformational changes in the Sdp1 active site. Apo (pink), sulphate- (gold) and phosphotyrosine-bound (blue) Sdp1. Right panel shows a superposition of Sdp (phosphotyrosine-bound) and VHR (magenta), indicating His 111 is structurally equivalent to aspartic acid 92 of VHR. **b**, Stereoview of a SIGMAA-weighted ($2F_o$–F_c) electron density map contoured at 1σ close to the Sdp1 active site, highlighting density for the His 111 side chain and Cys 47–Cys 142 disulphide bridge. The phosphotyrosine ring is shown in green and red, mainchain atoms for phosphotyrosine omitted from the refined model are shown in grey. **c**, Surface representation of the active sites of Sdp1, VHR and PTP1B with phosphotyrosine substrates, and equivalent phosphotyrosine contact residues, identified.

His 111 is structurally equivalent to the general acid aspartic acid residue of VHR (Fig. 3a, right panel). However, although it is required for ERK2 dephosphorylation it does not have a catalytic role (see Supplementary Material and Supplementary Fig. 2) and is likely to assist in phosphotyrosine recognition. His 111 is part of a loop motif Trp-X-His-X-Ser/Thr in which the Trp is buried and the Ser side chain stabilizes the loop conformation (Supplementary Fig. 1b). Sequence searches with the Trp-X-His-X-Ser/Thr and Cys-$(X)_5$-Arg motifs found matches exclusively within yeast and filamentous fungi genomes (Fig. 4a). We found two hits with these motifs in *Saccharomyces cerevisiae* corresponding to Sdp1 (YIL113W; *Saccharomyces* Genome Database at http://www.yeastgenome.org) and Msg5 (YNL053L). Both phosphatases are known to target Slt2 and their respective genes map to regions related by whole-genome duplication[19,26,27] (Fig. 4a). To test whether Msg5 is also sensitive to reducing agents, we assayed the activity of Msg5 towards pNPP and found that, like Sdp1, it exhibited a much poorer K_m in the presence of DTT than in its absence (Fig. 4b). Similarly, we found that mutation of the two cysteines C80S/C321S, analogous to the Sdp1 double cysteine mutant, eliminated the sensitivity to DTT (Fig. 4b). Thus, Sdp1 and Msg5 have equivalent catalytic properties.

Budding yeasts (hemi-ascomycetes) that phylogenetically branched off before the *S. cerevisiae* genome duplication, as well as filamentous fungi and fission yeast genomes, encode a single phosphatase bearing the Trp-X-His-X-Ser/Thr and C-$(X)_5$-Arg motifs (Fig. 4a). All budding yeast hits with these motifs have retained an equivalent cysteine to Cys 142 of Sdp1, suggesting that they too will be regulated in a similar manner to Sdp1 and Msg5 (Fig. 4a). The precise half-cysteine partner for Cys 142 cannot be reliably predicted in these cases owing to lack of sequence conservation and the presence of multiple cysteine residues in their respective N-terminal domains (data not shown). In contrast, matches to phosphatases encoded by filamentous fungi and fission yeast genomes lack an equivalent Cys 142, suggesting they may exhibit a different regulatory mechanism from Sdp1, Msg5 and budding yeast equivalents. For example, Pmp1 from fission yeast lacks a Cys 142 equivalent and requires the presence of DTT for activity (Fig. 4a and 4b). It has a methionine at this position (Fig. 4a), suggesting that this residue could functionally substitute for the intramolecular disulphide bridge. Consistent with this idea, we found that a C142M mutant Sdp1 retains some activity towards ERK2 but lost any sensitivity towards DTT (Fig. 2c and data not shown).

Our results establish a new group of MAPK phosphatases specific for yeast and fungal taxa that we propose should be classified as WH phosphatases because the Trp and His residues within the

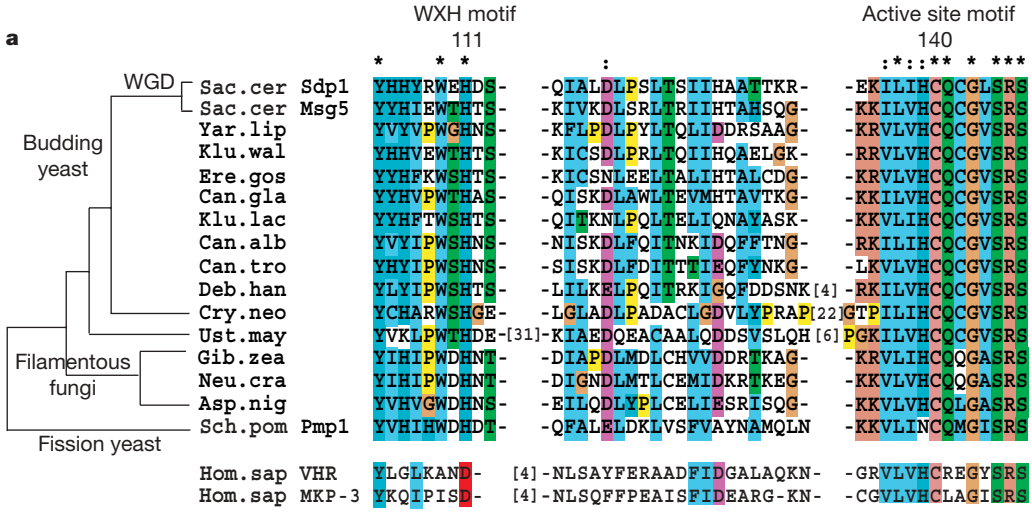


Figure 4 | Sdp1 and Msg5 are prototypic members of the WH phosphatase family. a, Sequence alignment of Sdp1 and related WH phosphatases close to their active site. Structurally equivalent sequences from MKP-3 and VHR are shown on the bottom for comparison. Selected organisms with a WH phosphatase are clustered according to consensus phylogeny indicating hemiascomycotina (budding yeasts), filamentous ascomycota (filamentous fungi) and representatives of basidiomycota (*Ustilago maydis*) and archiascomycotina (fission yeast) subphyla. Connectors are of arbitrary length. WGD, whole genome duplication. Sac cer, *Saccharomyces cerevisiae*; Sch pom, *Schizosaccharomyces pombe*; Yar lip, *Yarrowica lipolytica*; Klu wal, *Kluyveromyces waltii*; Ere gos, *Eremothecium gossypii*; Can gla, *Candida glabrata*; Klu lac, *Kluyveromyces lactis*; Can alb, *Candida albicans*; Can tro, *Candida tropicalis*; Deb han, *Debaryomyces hansenii*; Cry neo, *Cryptococcus neoformans*; Ust may, *Ustilago maydis*; Gib zea, *Gibberella zeae*; Neu cra, *Neurospora crassa*; Asp nig, *Aspergillus niger*; and Hom sap, *Homo sapiens*. **b**, Msg5 activity towards pNPP shows similar sensitivity to reducing agents as Sdp1. Equivalent conditions to those described in Fig. 1a were used with or without 1 mM DTT. Pmp1, which lacks a predicted intramolecular disulphide, does not exhibit reductant sensitivity. **c**, Model for the activation of Sdp1 under oxidative conditions leading to Slt2 dephosphorylation. *SDP1* synthesis is controlled by the Msn2/4 transcription factor (reviewed in ref. 13).

Trp-X-His-X-Ser/Thr motif are absolutely conserved in all cases (Fig. 4a). These phosphatases were not previously recognized as being closely related owing to sequence alignment errors close to the Trp-X-His-X-Ser/Thr motif (see DSPc/PF00782 Pfam entry at http://www.sanger.ac.uk/Software/Pfam). In light of our findings, published biochemical experiments may need to be re-evaluated for Sdp1 and Msg5. They also raise the question as to whether there are functionally analogous phosphatases in higher eukaryotes operating under oxidative conditions that would inactivate MAPKs[28]. Phosphatase oxidation has emerged as an important control mechanism to regulate these enzymes negatively in response to cell surface receptor activation[4,10]. Our observations identify an unprecedented mechanism for protein tyrosine phosphatase activation through a disulphide bridge that does not involve the catalytic cysteine (Fig. 4c). Once redox equilibrium is restored inside the cell, Sdp1 activation would most probably be reversed by reduction of this disulphide bond. Slt2 and Sdp1 synthesis is also responsive to heat shock and this stress may increase oxidative metabolism leading to higher levels of endogenous reactive oxygen species that in turn activate Sdp1. Yeast employ a variety of mechanisms to recognize and adapt to oxidative stress that frequently involve the reversible formation of disulphide bonds[29,30]. The existence of an analogous mechanism for Sdp1 shows that yeast have evolved a strategy to overcome the intrinsic limitation of using thiol-based phosphatases to counter and reverse oxidative-activation of a stress responsive MAPK.

METHODS

All proteins used in this study were produced in *Escherichia coli* and purified using an affinity tag. Crystals were grown using the hanging-drop vapour diffusion method. Diffraction data were collected at the European Synchrotron Radiation Facility at Grenoble and the Synchrotron Radiation Source at Daresbury. The apo/sulphate crystal structure was solved by molecular replacement using a composite search model of MKP-3 catalytic domain (Protein Data Bank, 1MKP) and VHR structures (PDB, 1VHR). The Sdp1(C140S)–phosphotyrosine structure was solved using the Sdp1/sulphate coordinates. Enzymatic assays were performed as previously described using either pNPP or recombinant diphosphorylated ERK2 as a substrate[20]. The free

sulphydryl content was measured using Ellman's reagent (Pierce Chemicals) according to the manufacturer's instructions. Tryptic digests of the Sdp1(C140S) mutant were analysed by mass spectrometry using an Applied Biosystems 4700 Proteomics Analyser. *In vivo* experiments used an Sdp1 null yeast strain transformed with a low copy number plasmid expressing a Myc-tagged Sdp1 or Sdp1 mutants driven by the endogenous *SDP1* promoter. After induction of Sdp1, yeast were exposed to H_2O_2, pelleted and lysed to produce protein extracts for analysis as described previously[23].

Received 13 February; accepted 3 April 2007.
Published online 9 May 2007.

1. Ikner, A. & Shiozaki, K. Yeast signaling pathways in the oxidative stress response. *Mutat. Res.* **569,** 13–27 (2005).
2. Chang, L. & Karin, M. Mammalian MAP kinase signalling cascades. *Nature* **410,** 37–40 (2001).
3. Keyse, S. M. Protein phosphatases and the regulation of mitogen-activated protein kinase signalling. *Curr. Opin. Cell Biol.* **12,** 186–192 (2000).
4. Tonks, N. K. Redox redux: revisiting PTPs and the control of cell signaling. *Cell* **121,** 667–670 (2005).
5. Salmeen, A. & Barford, D. Functions and mechanisms of redox regulation of cysteine-based phosphatases. *Antioxid. Redox Signal.* **7,** 560–577 (2005).
6. Barford, D., Das, A. K. & Egloff, M. P. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 133–164 (1998).
7. Alonso, A. *et al.* Protein tyrosine phosphatases in the human genome. *Cell* **117,** 699–711 (2004).
8. Pulido, R., Zuniga, A. & Ullrich, A. PTP-SL and STEP protein tyrosine phosphatases regulate the activation of the extracellular signal-regulated kinases ERK1 and ERK2 by association through a kinase interaction motif. *EMBO J.* **17,** 7337–7350 (1998).
9. Mattison, C. P., Spencer, S. S., Kresge, K. A., Lee, J. & Ota, I. M. Differential regulation of the cell wall integrity mitogen-activated protein kinase pathway in budding yeast by the protein tyrosine phosphatases Ptp2 and Ptp3. *Mol. Cell. Biol.* **19,** 7651–7660 (1999).
10. Rhee, S. G., Bae, Y. S., Lee, S. R. & Kwon, J. Hydrogen peroxide: a key messenger that modulates protein phosphorylation through cysteine oxidation. *Sci. STKE* **1,** doi:10.1126/stke.2000.53.pe1 (2000).
11. Salmeen, A. *et al.* Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate. *Nature* **423,** 769–773 (2003).
12. Seth, D. & Rudolph, J. Redox regulation of MAP kinase phosphatase 3. *Biochemistry.* 45,–8476–8487 (2006).
13. Martin, H., Flandez, M., Nombela, C. & Molina, M. Protein phosphatases in MAPK signalling: we keep learning from yeast. *Mol. Microbiol.* **58,** 6–16 (2005).
14. Staleva, L., Hall, A. & Orlow, S. J. Oxidative stress activates *FUS1* and *RLM1* transcription in the yeast *Saccharomyces cerevisiae* in an oxidant-dependent manner. *Mol. Biol. Cell* **15,** 5574–5582 (2004).
15. Alic, N., Higgins, V. J., Pichova, A., Breitenbach, M. & Dawes, I. W. Lipid hydroperoxides activate the mitogen-activated protein kinase Mpk1p in *Saccharomyces cerevisiae. J. Biol. Chem.* **278,** 41849–41855 (2003).
16. Vilella, F., Herrero, E., Torres, J. & de la Torre-Ruiz, M. A. Pkc1 and the upstream elements of the cell integrity pathway in *Saccharomyces cerevisiae*, Rom2 and Mtl1, are required for cellular responses to oxidative stress. *J. Biol. Chem.* **280,** 9149–9159 (2005).
17. Krasley, E., Cooper, K. F., Mallory, M. J., Dunbrack, R. & Strich, R. Regulation of the oxidative stress response through Slt2p-dependent destruction of cyclin C in *Saccharomyces cerevisiae. Genetics* **172,** 1477–1486 (2006).
18. Levin, D. E. Cell wall integrity signaling in *Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev.* **69,** 262–291 (2005).
19. Hahn, J. S. & Thiele, D. J. Regulation of the *Saccharomyces cerevisiae* Slt2 kinase pathway by the stress-inducible Sdp1 dual specificity phosphatase. *J. Biol. Chem.* **277,** 21278–21284 (2002).
20. Collister, M. *et al.* YIL113w encodes a functional dual-specificity protein phosphatase which specifically interacts with and inactivates the Slt2/Mpk1p MAP kinase in *S. cerevisiae. FEBS Lett.* **527,** 186–192 (2002).
21. Groom, L. A., Sneddon, A. A., Alessi, D. R., Dowd, S. & Keyse, S. M. Differential regulation of the MAP, SAP and RK/p38 kinases by Pyst1, a novel cytosolic dual-specificity phosphatase. *EMBO J.* **15,** 3621–3632 (1996).
22. Todd, J. L., Tanner, K. G. & Denu, J. M. Extracellular regulated kinases (ERK) 1 and ERK2 are authentic substrates for the dual-specificity protein-tyrosine phosphatase VHR. A novel role in down-regulating the ERK pathway. *J. Biol. Chem.* **274,** 13271–13280 (1999).
23. Delauney, A., Pflieger, D., Barrault, M. B., Vinh, J. & Toledano, M. B. A thiol peroxidase is an H_2O_2 receptor and redox-transducer in gene activation. *Cell* **111,** 471–481 (2002).
24. Jia, Z., Barford, D., Flint, A. J. & Tonks, N. K. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science* **268,** 1754–1758 (1995).
25. Stewart, A. E., Dowd, S., Keyse, S. M. & McDonald, N. Q. Crystal structure of the MAPK phosphatase Pyst1 catalytic domain and implications for regulated activation. *Nature Struct. Biol.* **6,** 174–181 (1999).
26. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae. Nature* **428,** 617–624 (2004).
27. Flandez, M., Cosano, I. C., Nombela, C., Martin, H. & Molina, M. Reciprocal regulation between Slt2 MAPK and isoforms of Msg5 dual-specificity protein phosphatase modulates the yeast cell integrity pathway. *J. Biol. Chem.* **279,** 11027–11034 (2004).
28. Kamata, H. *et al.* Reactive oxygen species promote TNFα-induced death and sustained JNK activation by inhibiting MAP kinase phosphatases. *Cell* **120,** 649–661 (2005).
29. Barford, D. The role of cysteine residues as redox-sensitive regulatory switches. *Curr. Opin. Struct. Biol.* **14,** 679–686 (2004).
30. Veal, E. A. *et al.* A 2-Cys peroxiredoxin regulates peroxide-induced oxidation and activation of a stress-activated MAP kinase. *Mol. Cell* **15,** 129–139 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements G.C.F. was supported by a studentship from GlaxoSmithKline. N.Q.M. and S.M.K. are funded by CR-UK. M.J.R.S. is supported by BBSRC. We thank M. Toledano for experimental advice, T. Kuno for the Pmp1 plasmid and M. Way for critical comments on the manuscript. We also thank M. Cobb for the gift of the ERKpTpY expression vector, D. Thiele and J.-S. Hahn for yeast strains and Sdp1 expression plasmids, and A. Stewart, S. Tilley and J. Denu for assistance with early experiments.

Author Contributions G.C.F. and N.Q.M. determined the apo and sulphate-bound structures. M.S. determined the phosphotyrosine structure and carried out pNPP and ERK2 assays and blots. D.C.B. and N.Q.M. refined the structures. P.P.K. measured the free -SH content and reproduced all kinetic experiments. S.M.K. and M.J.R.S. designed the *in vivo* assays. M.C, M.J.D, V.M., and R.J.D. carried out *in vivo* assays using Myc-tagged Sdp1 proteins. S.H. and N.T carried out the mass spectrometry. S.M.K. and N.Q.M planned the project and designed the experiments. N.Q.M., S.M.K. and D.C.B. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Coordinates have been deposited at the PDB with accession codes 2j16 (apo/sulphate-bound Sdp1) and 2j17 (phosphotyrosine–Sdp1 complex). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.Q.M. (neil.mcdonald@cancer.org.uk).

LETTERS

Experimental and theoretical study of mitotic spindle orientation

Manuel Théry[1,2]*, Andrea Jiménez-Dalmaroni[3]*, Victor Racine[1], Michel Bornens[1] & Frank Jülicher[3]

The architecture and adhesiveness of a cell microenvironment is a critical factor for the regulation of spindle orientation *in vivo*[1,2]. Using a combination of theory and experiments, we have investigated spindle orientation in HeLa (human) cells. Here we show that spindle orientation can be understood as the result of the action of cortical force generators, which interact with spindle microtubules and are activated by cortical cues. We develop a simple physical description of this spindle mechanics, which allows us to calculate angular profiles of the torque acting on the spindle, as well as the angular distribution of spindle orientations. Our model accounts for the preferred spindle orientation and the shape of the full angular distribution of spindle orientations observed in a large variety of different cellular microenvironment geometries. It also correctly describes asymmetric spindle orientations, which are observed for certain distributions of cortical cues. We conclude that, on the basis of a few simple assumptions, we can provide a quantitative description of the spindle orientation of adherent cells.

The control of mitotic spindle orientation in classical developmental systems is mainly based on the activity of cortical cues[2–6]. These cues can either be intrinsic, due to cell polarity, or extrinsic, such as cues associated with the cell's contacts to its microenvironment[7–9]. Recently, it was shown that HeLa cells that divide on fibronectin-coated micropatterns orient their spindle relative to the pattern geometry[10]. During division, HeLa cells round up but remain attached to the adhesive pattern by retraction fibres[10,11] (Fig. 1). Some actin-associated proteins accumulate in the cell cortex at the end of these fibres. They constitute cortical cues that are possibly implicated in spindle orientation[6,10]. In the present work, we study both experimentally and theoretically the interplay of these cortical cues and spindle mechanics that governs spindle orientation. Cortical force generators pull on astral microtubules radiating from spindle poles[3,4,12,13]. This results in a net torque on the spindle that induces its rotation. We show that a simple physical description of this spindle mechanics can quantitatively account for the observed distribution of spindle orientations.

The central idea of our theoretical approach is that cortical force generators are locally activated by cortical cues that are associated with the adhesive microenvironment of the cell. In the case of cells in culture, these cortical cues are correlated with the presence of retraction fibres[6,10]. More specifically, we assume that the cortical force exerted per microtubule acts in a direction tangential to the

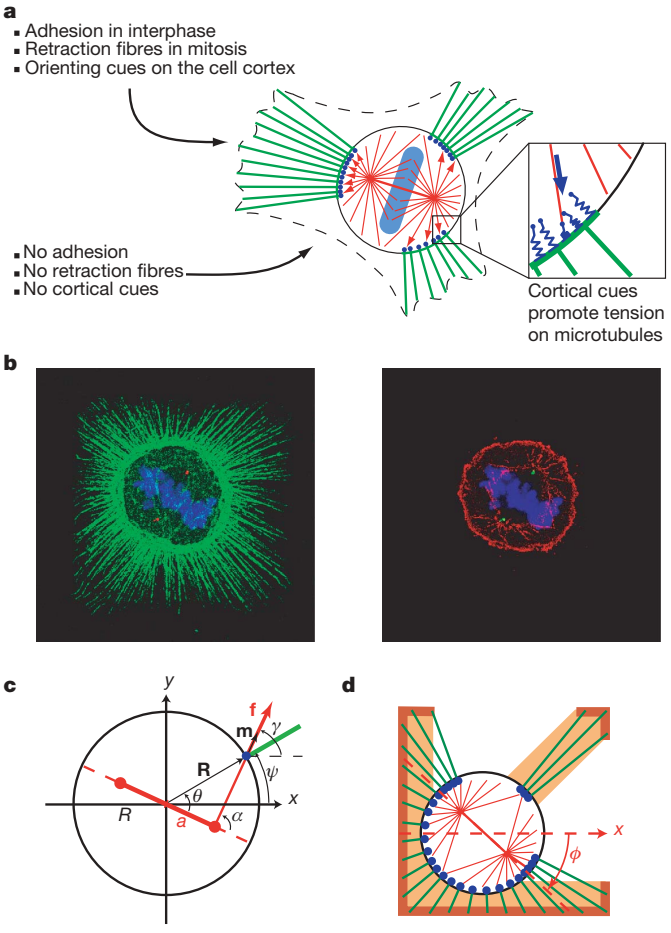


Figure 1 | Mitotic spindle orientation and cortical forces. **a**, Schematic representation of a spherical cell during mitosis (circle) linked by retraction fibres (green) to adhesion sites. Our key assumption is that the density of retraction fibres at the cortex activates cortical force generators (blue), which exert pulling forces on astral microtubules. As a result, a torque acts on the mitotic spindle (red), which rotates it as well as the metaphase plate (cyan) until a stable orientation angle is attained. **b**, Mitotic cell in metaphase on a frame-shaped micropattern after fixation with glutaraldehyde. Left: actin (green), spindle poles (red) and chromosomes (blue). Right: astral microtubules (red), spindle poles (green) and chromosomes (blue). **c**, Schematic representation of spindle geometry and cortical forces. Spindle poles (red) are separated by a distance $2a$ in a cell of radius R. Cortical force generators exert a pulling force **f** tangential to the orientation of astral microtubules described by the unit vector **m**. This force exerts a torque $\mathbf{R} \times \mathbf{f}$ on the spindle. Here, **R** is the vector pointing to the cortical position at which the force acts. We used the values $R = 10\ \mu m$ and $a = 6\ \mu m$, determined experimentally. **d**, Geometry of a cell on an arrow-shaped adhesive pattern (orange) during mitosis. The dark orange outline corresponds to the pattern edge along which retraction fibres attach. This outline is given by the convex hull of the pattern. The orientation of the spindle is described by the angle ϕ between the spindle axis and an horizontal reference line.

[1]Institut Curie, CNRS UMR144, Compartimentation et Dynamique Cellulaire, 26 rue d'Ulm, 75248 Paris, France. [2]Commissariat à l'Energie Atomique, DSV, iRTSV, Laboratoire Biopuces, 17 rue des Martyrs, 38054 Grenoble, France. [3]Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany.
*These authors contributed equally to this work.

microtubule[3,12] and is proportional to the local density of retraction fibres ending at the cell cortex (Fig. 1a). Using this assumption, we calculate the torques that govern spindle rotation, provided that the geometry of the spindle and the spatial distribution of retraction fibres are known. We use a simplified two-dimensional geometric representation of the spindle, which neglects microtubule bending and spindle deformations. Furthermore, we use the following assumptions, which are based on our experimental observations: (1) cells round up during division and attain a spherical shape (Fig. 1b); (2) retraction fibres emerge radially from the spherical cell body and extend to the convex hull of the adhesive pattern (Fig. 1b); (3) the density of retraction fibres along the pattern outline is constant (Fig. 1b); (4) the cell centre during division is located near the centre of mass of the pattern (Supplementary Fig. S1); and (5) displacement of the spindle away from the cell centre can be neglected (Supplementary Fig. S2).

The cortical force exerted per angular element on the spindle in a direction tangential to the astral microtubules is then given by:

$$\mathbf{f}(\psi,\phi) = F(\psi)\rho_{\mathrm{MT}}(\psi-\phi)\mathbf{m}(\psi,\phi) \quad (1)$$

Here $F(\psi)$ denotes the magnitude of the force acting per microtubule at the cortical angle ψ. The unit vector that points outwards in the direction of the astral microtubules at cortical angle ψ if the spindle axis is at an angle ϕ relative to a reference axis (Fig. 1c) is denoted $\mathbf{m}(\psi,\phi)$. The angular density of microtubules $\rho_{\mathrm{MT}}(\theta)$ reaching the cell cortex at angle $\theta = \psi - \phi$ relative to the spindle axis depends on the total number N_{MT} of microtubules that emerge from one spindle pole in the planar projection and on the spindle geometry (see Supplementary Information).

Our key assumption, that retraction fibres locally activate cortical force generators, implies that the force $F(\psi)$ is proportional to $\rho_{\mathrm{r}}(\psi)$, where $\rho_{\mathrm{r}}(\psi)$ denotes the angular density of retraction fibres reaching the cortex at angle ψ. The strength of this coupling of retraction fibre density to motor activation is characterized by the coefficient C, which has units of force (see Supplementary Information). The density ρ_{r} can be estimated using the pattern geometry (see Fig. 1d and Supplementary Information). The net torque exerted on the spindle is given by the vector

$$\boldsymbol{\tau}(\phi) = \int_{-\pi}^{\pi} \mathrm{d}\psi\, \mathbf{R}\times\mathbf{f} \quad (2)$$

where the vector $\mathbf{R}(\psi)$ points from the cell centre to the cortical position with angle ψ (Fig. 1c). The torque $\boldsymbol{\tau}$ depends only on the spindle orientation ϕ. It is convenient to define the effective energy landscape

$$W(\phi) = -\int_{-\pi/2}^{\phi} \tau_z(\phi')\mathrm{d}\phi' \quad (3)$$

where τ_z is the component of $\boldsymbol{\tau}$ normal to the x–y plane. Stable spindle orientations correspond to minima of the potential $W(\phi)$. Taking into account fluctuations as additional random torques with zero average, the system exhibits a distribution of spindle orientations that depends on the strength D of fluctuations. In our simple model, we find that the angular distribution of spindle orientations is of the form (see Supplementary Information)

$$P(\phi) = N\exp(-w(\phi)/d) \quad (4)$$

where N is a normalization factor, $w(\phi) = [2\pi/(CN_{\mathrm{MT}}R)]\,W(\phi)$ is a dimensionless energy landscape, R is the cell radius and $d = (2\pi D)/(CN_{\mathrm{MT}}R)$ is a dimensionless coefficient, which combines the effects of the noise strength and the strength of the coupling of retraction fibres to the activity of force generators as well as the numbers of force generators and microtubules.

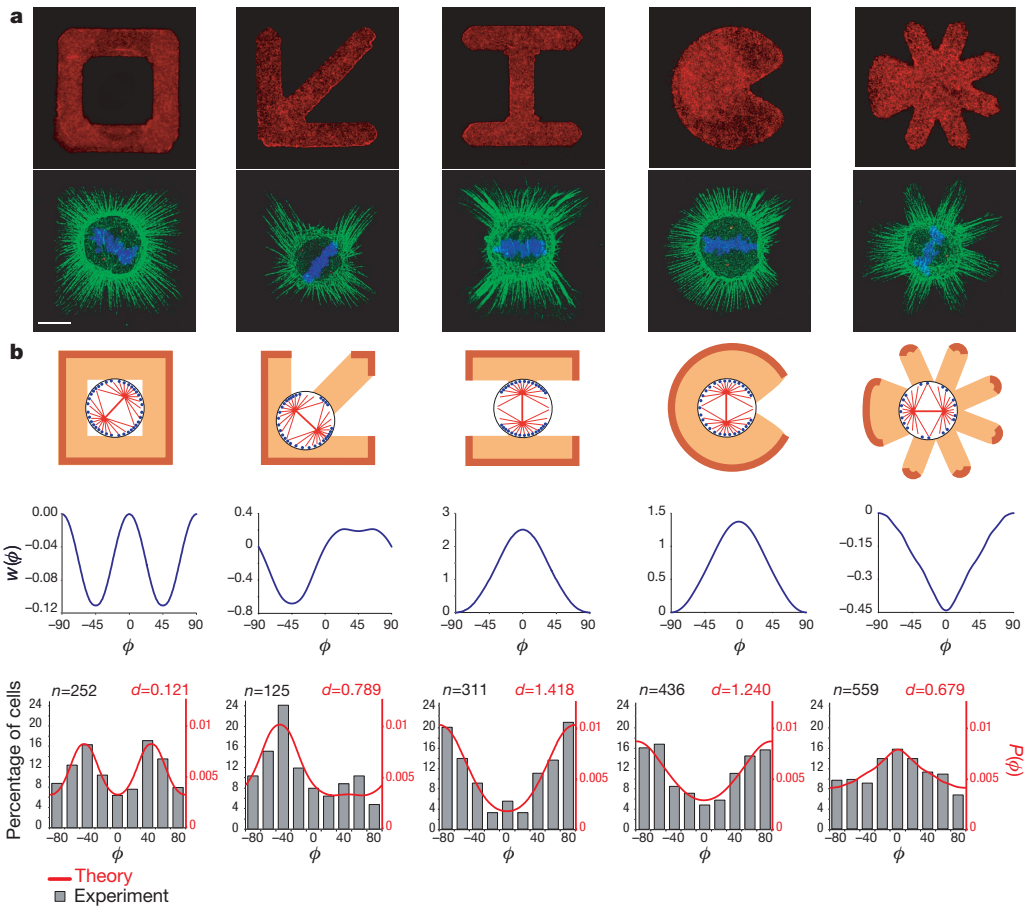


Figure 2 | Spindle orientation on adhesive micropatterns, showing theoretical results and experimental data. **a**, Mitotic cells on different fibronectin micropatterns (red, first row). Retraction fibres (green, second row), spindle poles (red) and the chromosomes (blue) were labelled. Scale bar, 10 μm. **b**, Top row, schematic representation of micro-patterns (orange), the zones of anchorage of retraction fibres (dark orange outline), and of the corresponding distributions of force generators (blue dots). The theoretical potential energy landscape $w(\phi)$ (blue, middle row) and the angular probability density of spindle orientation $P(\phi)$ (red, bottom row) were calculated. A single fit parameter d was used to fit the experimentally measured histograms of spindle orientations (grey, n measures on each pattern; bottom row).

The orientation of the mitotic spindle in HeLa cells was studied experimentally on various fibronectin micropatterns for which we measured the angular distributions of spindle orientations during division as well as the positions of cell centres (Methods). The experimentally determined cell centres were always located very close to the centre of mass of the patterns (Supplementary Fig. S1). Figure 2 shows the experimental and theoretical results obtained for five patterns. On the first three patterns—frame, arrow and H-shaped micropatterns—cells attained similar square shapes before division (not shown), but exhibited different distributions of retraction fibre densities and spindle orientations (Fig. 2a). Similarly, the two remaining patterns imposed a circular cell shape before division, but different distributions of retraction fibres and spindle orientations. For each pattern shape, we calculated the dimensionless energy landscape $w(\phi)$ describing the torques acting on the spindle. To compare experiments and theory, we used a fit of the calculated normalized angular distribution functions $P(\phi) = N \exp(-w(\phi)/d)$ to the experimental data using d as a single fit parameter. We found that our theory could correctly describe the most probable spindle orientation angle and furthermore could quantitatively account for the full shape of the angular distributions of spindle orientation (Fig. 2b). We also compared our theory with experimental data reported previously for a large set of different patterns (Supplementary Fig. S3). Our results show that for all these patterns the theory correctly describes the most probable orientation angle and the full angular distributions of orientations.

We then studied situations where variations in the pattern geometry lead to transitions in the preferred orientation angle. Using H-shaped patterns, we quantified the distributions of spindle orientations when the aspect ratio of the pattern was changed progressively. Such pattern transformations induce the displacement of the points where cortical cues are most pronounced (Fig. 3a). The most probable orientation angle observed experimentally changed from $\phi = 90^\circ$ to $\phi = 0^\circ$ (Fig. 3a). Intermediate patterns exhibited angular distributions with two maxima. Our model exhibited the same features when the aspect ratio of the pattern was varied. It could account for the observed changes of the angular distributions as a result of small displacements of the distributions of cortical cues (Fig. 3b). In most examples discussed so far, the most probable spindle orientation was symmetric, that is, normal to a symmetry axis of the pattern. However, the star-shaped pattern (Fig. 2, right pattern) interestingly leads to spindle orientation such that the cell division is oriented asymmetrically, that is, along the symmetry axis of the pattern. We investigated whether the preferred spindle orientation could switch from symmetric to asymmetric by a deformation of the pattern shape. Such a transition could indeed be observed (Fig. 4). Cells plated on arrow-shaped patterns divide mostly perpendicular with respect to the pattern axis, which corresponds to a symmetric orientation. If this pattern is deformed to a crossbow-shape, the preferred orientation is along the symmetry axis corresponding to asymmetric orientation (Fig. 4a). Our model can account for this transition from symmetric to asymmetric spindle orientation (Fig. 4b). This shows that by changing the strength of cortical cues at fixed cortical locations, the relative depth of potential minima can be affected. This can result in sharp transitions of the preferred orientation angles between closely related patterns (Fig. 4c).

Our physical description of spindle mechanics could account quantitatively for the observed distribution of spindle orientation angles on a large variety of different geometries of adhesive patterns (Figs 2–4 and Supplementary Fig. S3). Small quantitative differences between calculated and observed distributions of spindle orientations reveal that our simplified description does not capture all details of the experiments (Supplementary Information). Our results highlight the possibility that a slight modification of the cell microenvironment is sufficient to provide distinct signals that can induce a transition from symmetrical to asymmetrical orientation of the spindle, and consequently lead to an unequal cell division[2] (Fig. 4c). However, at this stage we cannot conclude that HeLa cells plated on star- or crossbow-shaped patterns undergo genuine asymmetric cell division (with unequal distribution of determinants between the two daughter cells), as these somatic non-stem cells do not express and segregate differentiation determinants.

These results demonstrate that cortical cues which have a distribution that is set up by the geometry of the adhesive pattern control spindle orientation by regulating torques that act on the spindle. Although we do not have direct evidence that the activity of cortical force generators is proportional to the local density of retraction fibres, the remarkable qualitative and also largely quantitative agreement between theory and experiment strongly supports this idea. This aspect will require direct molecular characterization[4,6,14–16]. It is noteworthy that the spindle mechanics discussed here—where rotations result from torques exerted by cortical force generators that are regulated by cortical cues—could be relevant to other systems. In tissues, cortical heterogeneity may not depend on retraction fibres but could still be set up by the microenvironment via the geometrical distribution of adhesion sites. Therefore our physical description of spindle orientation could prove useful in unravelling basic principles underlying tissue morphogenesis in developmental processes.

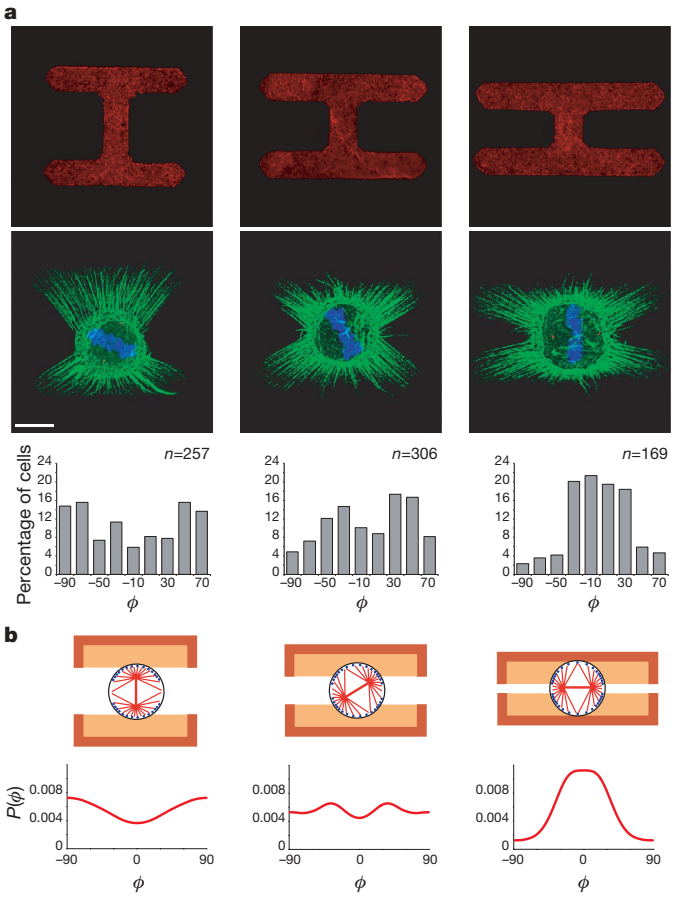


Figure 3 | Spindle orientation changes induced by continuous displacements of cortical cues. **a**, Mitotic cells on H-shaped patterns (red, first row) with varying aspect ratio. Retraction fibres (green, second row), spindle poles (red) and the chromosomes (blue) were labelled. Scale bar, 10 μm. Corresponding angular distributions of spindle orientation were measured (grey histograms). **b**, Three pattern shapes with different aspect ratios are shown (38, 44 and 50 μm wide, 34, 28 and 22 μm high from left to right) (orange) with corresponding distributions of force generators (blue dots). The theoretical potential energy landscape $w(\phi)$ (blue) and the angular probability density of spindle orientation $P(\phi)$ (red; bottom row) were calculated.

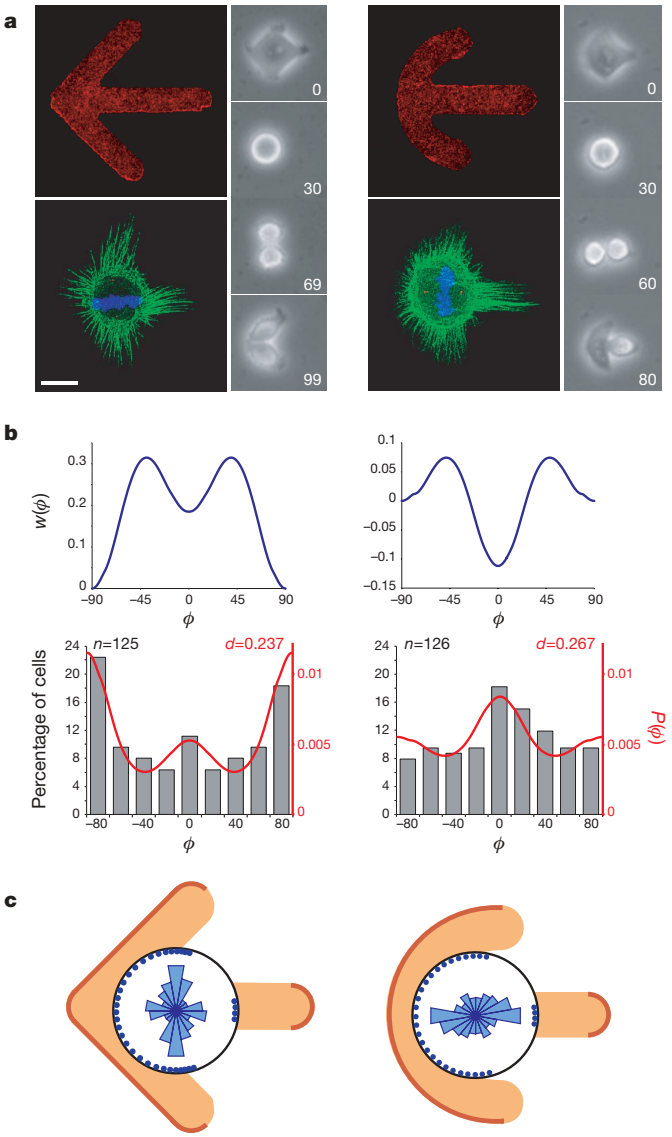


Figure 4 | Transition from symmetric to asymmetric spindle orientation. **a**, Mitotic cells on arrow-shaped (left) and on crossbow-shaped (right) fibronectin micropatterns (red). Retraction fibres (green, second row), spindle poles (red) and the chromosomes (blue) were labelled. Sequences of images acquired in time-lapse 10× phase contrast microscopy show examples of cell division on both patterns. Time is given in minutes. Scale bar, 10μm. **b**, The theoretical potential energy landscape $w(\phi)$ (blue) and the angular probability density of spindle orientation $P(\phi)$ (red) were calculated. A single fit parameter d was used to fit the experimentally measured histograms of spindle orientations (grey, n measures on each pattern). **c**, Pattern geometry used for the calculation of $w(\phi)$: the pattern outline along which retraction fibres are anchored (dark orange) and the corresponding density of force generators (blue dots) are indicated. The experimentally observed spindle orientation is displayed as angular histograms (light blue).

METHODS SUMMARY

The potential landscapes $w(\phi)$ were determined by numerically integrating equations (2) and (3) using the shape of a given pattern outline to determine the distribution of cortical cues $\rho_r(\psi)$ (Supplementary Information).

HeLa cells expressing centrin1–GFP were cultured, synchronized, plated on fibronectin micropatterns and video-recorded as previously described[10] (Methods). Briefly, fibronectin micropatterns were printed with a micro-structured polydimethylsiloxane stamp on a glass coverslip coated with mercaptosilane and passivated after microcontact printing with maleimide-polyethyleneglycol. HeLa cells were synchronized with a double thymidine block, trypsinized and plated on the printed glass coverslip and video-recorded at 37 °C with 10× magnification time-lapse phase contrast microscopy. The acquired cell division sequences were numerically processed for measurement of mitotic cell position, detection of anaphase and measurement of the orientation of cell elongation at this stage (Methods).

Cells were fixed with glutaraldehyde as previously described in order to preserve retraction fibres. Centrin–GFP decorates centrosomes at spindle poles. Actin was stained with phalloidin–Cy3 to reveal retraction fibres, and DNA was stained with Hoechst. Microtubules were immuno-labelled with mouse primary antibodies against β-tubulin (clone 2.1, 1/100, Sigma Aldrich) and anti-mouse Cy5-secondary antibodies (1/1000, Jackson Immunoresearch). Notably, a non-specific decoration of the cell cortex with these antibodies was induced by the glutaraldehyde fixation.

Image stacks of mitotic cells were acquired by performing $\acute{Z}$-acquisition with 500-nm steps from the cell bottom up to the middle of the cell using a Leica TCS-SP2 confocal microscope. Stacks were deconcolved and then projected to see in a single image the maximum value of each pixel of the stack.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 December 2006; accepted 2 April 2007.
Published online 9 May 2007.

1. Fuchs, E., Tumbar, T. & Guasch, G. Socializing with the neighbors: Stem cells and their niche. *Cell* **116,** 769–778 (2004).
2. Lu, B., Roegiers, F., Jan, L. Y. & Jan, Y. N. Adherens junctions inhibit asymmetric division in the *Drosophila* epithelium. *Nature* **409,** 522–525 (2001).
3. Grill, S. W., Howard, J., Schaffer, E., Stelzer, E. H. & Hyman, A. A. The distribution of active force generators controls mitotic spindle position. *Science* **301,** 518–521 (2003).
4. Colombo, K. *et al.* Translation of polarity cues into asymmetric spindle positioning in *Caenorhabditis elegans* embryos. *Science* **300,** 1957–1961 (2003).
5. Yamashita, Y. M., Jones, D. L. & Fuller, M. T. Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome. *Science* **301,** 1547–1550 (2003).
6. Thery, M. & Bornens, M. Cell shape and cell division. *Curr. Opin. Cell Biol.* **18,** 648–657 (2006).
7. Lechler, T. & Fuchs, E. Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* **437,** 275–280 (2005).
8. Gong, Y., Mo, C. & Fraser, S. E. Planar cell polarity signalling controls cell division orientation during zebrafish gastrulation. *Nature* **430,** 689–693 (2004).
9. Siegrist, S. E. & Doe, C. Q. Extrinsic cues orient the cell division axis in *Drosophila* embryonic neuroblasts. *Development* **133,** 529–536 (2006).
10. Thery, M. *et al.* The extracellular matrix guides the orientation of the cell division axis. *Nature Cell Biol.* **7,** 947–953 (2005).
11. Mitchison, T. J. Actin based motility on retraction fibers in mitotic PtK2 cells. *Cell Motil. Cytoskeleton* **22,** 135–151 (1992).
12. Grill, S. W., Kruse, K. & Julicher, F. Theory of mitotic spindle oscillations. *Phys. Rev. Lett.* **94,** 108104 (2005).
13. Pecreaux, J. *et al.* Spindle oscillations during asymmetric cell division require a threshold number of active cortical force generators. *Curr. Biol.* **16,** 2111–2122 (2006).
14. Izumi, Y., Ohta, N., Hisata, K., Raabe, T. & Matsuzaki, F. *Drosophila* Pins-binding protein Mud regulates spindle-polarity coupling and centrosome organization. *Nature Cell Biol.* **8,** 586–593 (2006).
15. Du, Q. & Macara, I. G. Mammalian Pins is a conformational switch that links NuMA to heterotrimeric G proteins. *Cell* **119,** 503–516 (2004).
16. Sanada, K. & Tsai, L. H. G protein βγ subunits and AGS3 control spindle orientation and asymmetric cell fate of cerebral cortical progenitors. *Cell* **122,** 119–131 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgments We thank A. Pépin and Y. Chen for technical help with micropattern fabrication, J.-B. Sibarita for technical help with video-microscopy, D. Grunwald for technical help with confocal image acquisitions and Y. Bellaïche for discussions.

Author Contributions M.T. performed experimental work, A.J.-D. performed numerical calculations, V.R. designed the software for movie analyses, and M.T., A.J.-D., M.B. and F.J. conceived the theoretical model.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.B. (michel.bornens@curie.fr) or F.J. (julicher@pks.mpg.de).

METHODS

Stamp fabrication. Pattern design was first done via L-Edit CAD software (Tanner EDA) and transferred to a machine-specific format corresponding to the electron beam lithography tool (Leica EBPG 5000 + nanowriter). Electron-beam lithography was then carried out on a blank 4 inch chromium-on-glass optical mask coated with resist (Nanofilm Inc). Resist development was done in pure AZ-Developer (Clariant) for 30 s. Chromium etch was then performed in chrome-etchant 3144 Puranal (Honeywell) for 1 min. The optical mask fabrication was completed after resist dissolution in acetone.

To make the resist mould, SPR220-7.0 photoresist (Shipley) was spin-coated at 2,000 r.p.m. for 1 min on a silicon wafer and soft-baked for 3 min at 115 °C resulting in a 9-μm-thick layer. Contact optical lithography was carried out using the fabricated optical mask in a Süss MA750 MicroTec mask aligner (UV source 405 nm, UV power 6 mW cm^{-2}) for 45 s. The photoresist was then developed for 2 min in pure LDD26W developer (Shipley). The obtained resist master mould was then exposed to chlorotrimethylsilane (Sigma-Aldrich) in the vapour phase, for PDMS anti-adhesion purposes.

PDMS (Sylgard 184 kit, Dow Corning) was finally cast on the resist mould and cured for 3 h at 60 °C. The 2-mm-thick cross-linked PDMS layer was peeled off, and stamps were manually cut out of it.

Microcontact printing. Glass coverslips were first washed in methanol/chloroform (50/50) for 24 h and stored in pure ethanol. After drying (15 min at 60 °C), a coverslip was oxidized in a plasma chamber (Harrick Plasma) for 3 min under a weak flow of air and incubated in a closed reactor containing a silanization mix of methanol, deionized water 4.5%, acetic acid 0.9%, 3 mercapto-propyltrimethoxy silane (S10475, Fluorochem) 2.5%, overnight at 4 °C (ref. 17). Coverslips were then washed twice in methanol and dried under filtered air followed by 15 min at 60 °C.

The PDMS stamp was oxidized in the plasma chamber for 10 s under a weak flow of air and inked with a 50 μg ml^{-1} fibronectin solution (Sigma-Aldrich) 10% of which was labelled with Cy3 (Amersham Biosciences) for 10 min. After aspiration of the fibronectin solution, the stamp was dried with filtered airflow and placed in contact with the silanized coverslip for 5 min. After removal of the stamp, the printed coverslip was immersed in a 20 mg ml^{-1} solution of poly(ethyleneglycol)-maleimide (2D2MOH01, Nektar Therapeutics) for 1 h at room temperature. The coverslip was then washed in PBS before cell deposition.

Cell culture and labelling. HeLa-B, human adenocarcinoma epithelial cell line, stably expressing centrin1–GFP (ref. 18), were cultured in DME medium with 10% fetal calf serum and 2 mM glutamine at 37 °C. Cells were synchronized at the G1–S transition using a double thymidine block and then removed from their flask using VERSEN, 10 min at 37 °C. After centrifugation, cells were resuspended in DMEM with 1% FCS and deposited on the printed coverslip at a density of 10^4 cells cm^{-2}.

Cells were fixed with glutaraldehyde as previously described[11] to preserve retraction fibres. Centrin–GFP decorates centrosomes at spindle poles. Actin was stained with phalloidin–Cy3 to reveal retraction fibres and DNA was stained with Hoechst. Microtubules were immuno-labelled with mouse primary antibodies against β-tubulin (clone 2.1, 1/100, Sigma Aldrich) and anti-mouse Cy5-secondary antibodies (1/1000, Jackson Immunoresearch). Notably, a non-specific decoration of the cell cortex with these antibodies was induced by the glutaraldehyde fixation.

Video microscopy. We used an inverted Leica DMIRBE microscope (Leica Microsystèmes) with a heated and motorized stage combined with a home-made plastic cell chamber to hold the printed glass coverslip, which was covered by a porous membrane allowing CO_2 buffering at pH 7.4. Metamorph software (Universal Imaging) was used for image acquisition. Hundreds of divisions were recorded in a few hours using time-lapse phase contrast microscopy on a multi-field acquisition at a frame rate of one picture every 3 min with a 10× magnitude objective.

Video analysis and processing. We developed software that was able to automatically recognize a single fluorescent micro-pattern within a field and detect the presence of a single cell attached to it[10]. Individual cell divisions were then extracted from the 10× phase contrast time-lapse recordings and every picture automatically segmented using a wavelet decomposition and fitted with an ellipse. The moment of cell elongation in anaphase was precisely detected, as the shape factor, defined as the ellipse length ratio, suddenly dropped from more than 0.9 to less than 0.6. The angle between the major axis of the ellipse, corresponding to spindle orientation, and a vertical reference was then recorded. The position of the centre of the round mitotic cell with respect to the pattern was also automatically recorded 6 min before anaphase.

17. Cuvelier, D., Rossier, O., Bassereau, P. & Nassoy, P. Micropatterned ''adherent/repellent'' glass surfaces for studying the spreading kinetics of individual red blood cells onto protein-decorated substrates. *Eur. Biophys. J.* **32**, 342–354 (2003).
18. Piel, M., Meyer, P., Khodjakov, A., Rieder, C. L. & Bornens, M. The respective contributions of the mother and daughter centrioles to centrosome activity and behavior in vertebrate cells. *J. Cell Biol.* **149**, 317–330 (2000).

LETTERS

Control of alternative RNA splicing and gene expression by eukaryotic riboswitches

Ming T. Cheah[1]*, Andreas Wachter[1]*, Narasimhan Sudarsan[1] & Ronald R. Breaker[1,2,3]

Bacteria make extensive use of riboswitches[1,2] to sense metabolites and control gene expression, and typically do so by modulating premature transcription termination or translation initiation. The most widespread riboswitch class known in bacteria responds to the coenzyme thiamine pyrophosphate (TPP)[3,4], which is a derivative of vitamin B_1. Representatives of this class have also been identified[5,6] in fungi and plants, where they are predicted[5,7] to control messenger RNA splicing or processing. We examined three TPP riboswitches in the filamentous fungus *Neurospora crassa*, and found that one activates and two repress gene expression by controlling mRNA splicing. A detailed mechanism involving riboswitch-mediated base-pairing changes and alternative splicing control was elucidated for precursor *NMT1* mRNAs, which code for a protein involved in TPP metabolism. These results demonstrate that eukaryotic cells employ metabolite-binding RNAs to regulate RNA splicing events that are important for the control of key biochemical processes.

A previous study[7] of the fungus *Aspergillus oryzae thiA* mRNA carrying a TPP aptamer revealed that thiamine (vitamin B_1) supplementation of growth medium results in reduced gene expression, and that deletion of riboswitch aptamer portions disrupts responsiveness to thiamine. These findings suggest that thiamine enters cells, is phosphorylated to generate TPP, and the resulting coenzyme serves as a ligand for riboswitch-mediated control of RNA splicing in fungi[5]. In this study, we examined the functions of TPP aptamers (Supplementary Fig. 1) present in three genes in *N. crassa*. Two of these genes, *NMT1* (Fig. 1a) and *CyPBP37* (Fig. 1b; a homologue of *THI4* and termed such hereafter), are known to be thiamine metabolism genes[8–10]. A third gene, *NCU01977.1* (Fig. 1c), codes for a protein of unknown function.

We primarily focused our examination on the *NMT1* gene, which is known to be repressed by excess thiamine in *N. crassa*[8] and in *Schizosaccharomyces pombe*[11]. All three precursor mRNAs carry the riboswitch in an intron residing near the 5′ terminus (Fig. 1, Supplementary Fig. 2). We used reverse transcription and polymerase chain reaction (RT–PCR) methods to establish the relative amounts and the nucleotide sequences of the 5′ ends of the transcripts produced when *N. crassa* was grown in the absence or presence of thiamine (Fig. 1d). Cloning and sequencing of RNA products (data not shown) revealed the presence of precursor transcripts that matched the sequence of the genomic DNA, and of other sequences that are consistent with the RNA splicing products depicted in Fig. 1. The results confirm that thiamine causes alternative splicing of the *NMT1* and *THI4* precursor mRNAs, and causes an increase in splicing of the *NCU01977.1* precursor mRNA. Thiamine does not affect splicing of an RNA that does not carry the TPP riboswitch (Supplementary Fig. 3).

TPP binding by *NMT1* RNA constructs was confirmed by in-line probing[12] (Supplementary Figs 4 and 5). Nucleotides that exhibit major structural modulations are known to be involved in the binding of TPP[13–15], and the apparent dissociation constant (K_D) of ~300 pM is similar to those exhibited by bacterial TPP riboswitches[3,16]. Thiamine control of *NMT1* RNA alternative splicing was further investigated by using RT–PCR to establish transcript types and amounts isolated from *N. crassa* grown in minimal medium and sampled at various times after thiamine supplementation

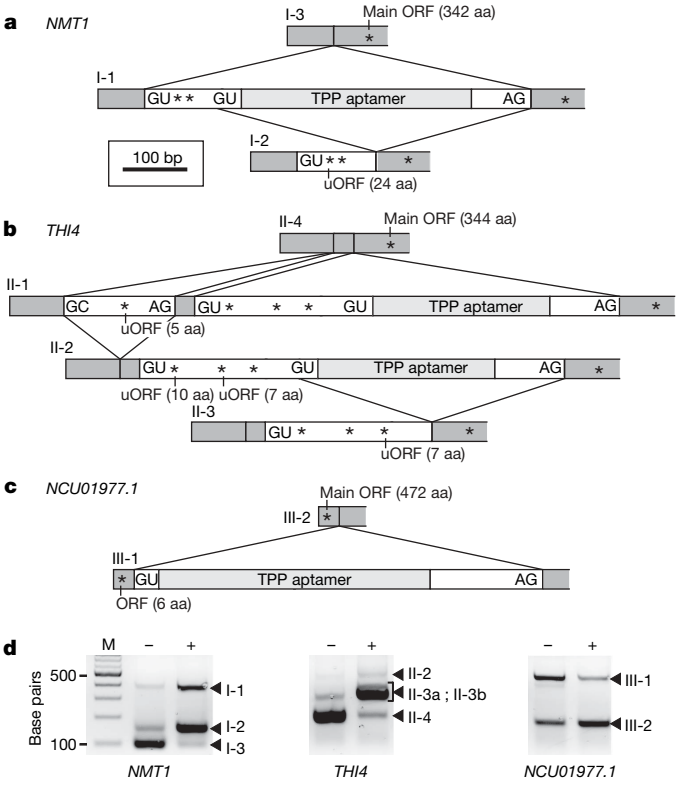


Figure 1 | Three *N. crassa* genes carry TPP riboswitches in 5′ introns. **a**, Precursor 5′ UTR (I-1) and alternatively spliced products (I-2 and I-3) for the *NMT1* mRNA. Exons and introns are represented by dark grey and unshaded/light grey rectangles, respectively. 5′ (GU) and 3′ (AG) splice sites, putative start codons (*) and the corresponding translation products from the upstream ORF (uORF) and the main *NMT1* ORF are depicted. aa, amino acids. **b**, **c**, 5′ regions of precursor mRNAs and their spliced products for *THI4* and *NCU01977.1*, respectively. **d**, RT–PCR detection of mRNA 5′ regions from *N. crassa* grown in the absence (−) or presence (+) of 30 μM thiamine. Bands II-3a and II-3b for *THI4* represent the splice form II-3 with the upstream intron remaining (3a) or removed (3b). Marker DNAs (M) are in 100 base pair increments.

[1]Department of Molecular, Cellular and Developmental Biology; [2]Department of Molecular Biophysics and Biochemistry; [3]Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06520, USA.
*These authors contributed equally to this work.

(Fig. 2a). In the absence of added thiamine ($t = 0$ min), transcripts are processed to yield splicing product I-3. Within the first hour after thiamine supplementation, the splicing product I-2 and the unspliced *NMT1* precursor RNA I-1 appear. Within 4 hours, I-3 is almost completely replaced by I-2 and I-1. These results suggest that the precipitous decrease of I-3 after the addition of thiamine is responsible for decreased *NMT1* expression.

Constructs carrying the *NMT1* 5′ UTR or its variants (Fig. 2b) with the start codon of the main open reading frame (ORF) fused in frame with a luciferase (*LUC*) reporter gene were used to assess the importance of the TPP aptamer for gene control. Substantial repression of the wild-type *LUC* reporter construct occurs with *N. crassa* grown overnight in medium supplemented with 30 μM thiamine (Fig. 2c, upper panel). Moreover, the RT–PCR products derived from the reporter construct and the native *NMT1* mRNAs are equivalent (Fig. 2c, lower panel). Most mutant constructs exhibit a two- to fourfold increase in reporter activity compared to wild type when cells are grown in thiamine-free medium. Mutations that weaken TPP binding affinity possibly also eliminate the modest level of gene repression caused by synthesis of TPP in cells growing in minimal medium.

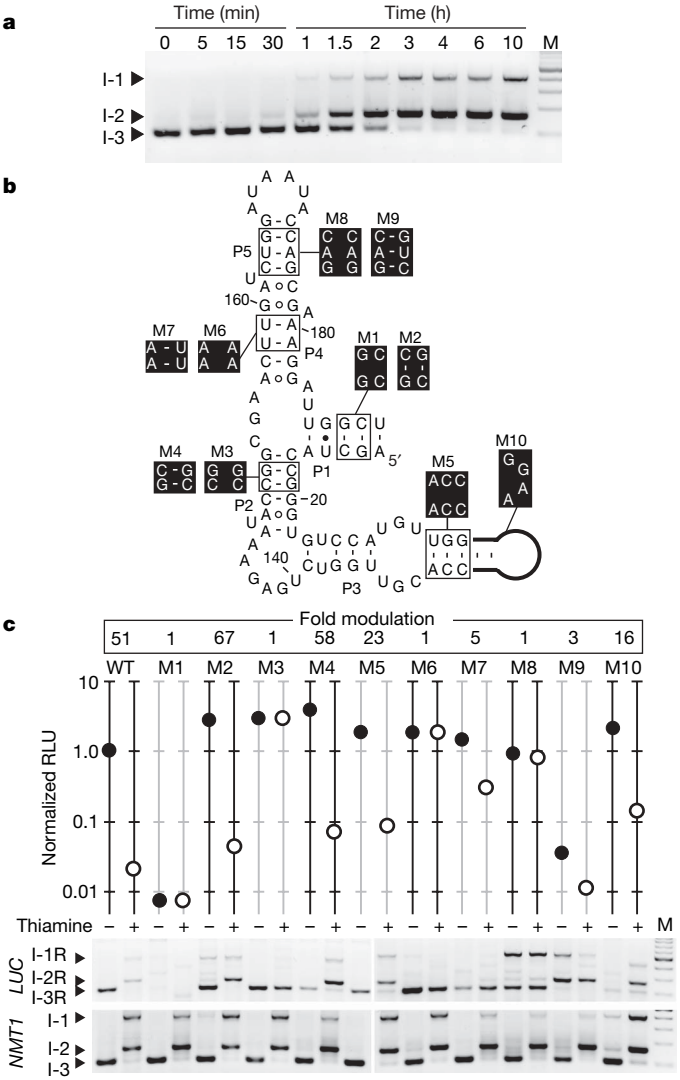


Figure 2 | Alternative splicing and gene control by the *NMT1* TPP riboswitch. a, Change in *NMT1* transcript splicing (RT–PCR products) after addition of 30 μM thiamine ($t = 0$) to a culture of *N. crassa* grown in thiamine-free medium. **b**, Reporter constructs (R) of wild-type (WT) or various mutant *NMT1* riboswitches (M1–M10) fused upstream of a luciferase (*LUC*) ORF. **c**, Upper panel, relative light units (RLU) from the WT–*LUC* construct (normalized RLU = 1) versus various mutant *NMT1–LUC* constructs grown in the absence (filled circles) or presence (open circles) of 30 μM thiamine. Values are the averages from three independent assay repeats and standard deviation error bars are smaller than the diameter of the symbols. Lower panel, RT–PCR analyses of the 5′ regions from the *LUC* fusions (upper panel) and the native *NMT1* RNA (lower panel) for each transformant. Details are described in Fig. 1 and Methods.

Mutations in stems P1 through P5 that disrupt and subsequently restore base pairing within the aptamer (constructs M1 through M9, Fig. 2b) mostly yield gene regulation characteristics that correlate with the ability of the RNAs to bind TPP. Although the extended portion of P3 is disrupted in M5, this change occurs outside the TPP-binding core of the aptamer and has little effect on gene control. In addition, the majority of the extended P3 stem can be deleted (M10, analogous to 115 *NMT1* RNA in Supplementary Fig. 4) without complete loss of function. Mutations in M1 cause a dramatic decrease of *LUC* activity and the typical mRNA products are not detected (Fig. 2c), indicating that some nucleotides and structures within the aptamer might influence mRNA transcription or processing in addition to their role in binding TPP.

The most revealing results were found with M9, which carries compensatory mutations to the disruptive mutations in the P5 stem of M8. M9 exhibits only partial restoration of thiamine-dependent gene control, but at a level of expression that is far below wild-type or other compensation mutants (Fig. 2c). These unusual characteristics of M9 are consistent with a mechanism whereby nucleotides within P5 participate in the control of alternative splicing (see below).

The effects of most mutations on thiamine regulation led us to speculate that unspliced or alternatively spliced RNAs are inactive owing to the presence of start codons upstream of the main ORF (Fig. 1a). This hypothesis was tested by examining additional *LUC* reporter constructs fused downstream of wild-type or mutant versions of the three types of *NMT1* 5′ UTRs (Fig. 3a). The resulting construct I-3R, which lacks upstream start codons and mimics the short (or fully) spliced mRNA and predominates in the absence of added thiamine, yields robust reporter activity (Fig. 3b). In contrast, the analogous I-2R construct exhibits almost no reporter activity, which is consistent with the natural production of this splice variant when thiamine is present and gene expression is reduced (Fig. 2c). The levels of *LUC* expression with constructs I-2R and I-3R are unchanged by the addition of thiamine, as expected as the TPP riboswitch is absent.

Disruption of the first (M11), second (M12) or both (M13) start codons in the alternatively spliced I-2R construct (Fig. 3a) upstream of the main *NMT1* ORF results in constructs that yield progressively more reporter expression. It has been observed that short upstream ORFs (uORFs) in the 5′ UTR of fungal genes decrease expression of the main ORF[17]. Therefore, restoration of *LUC* expression on disruption of both uORF start codons in I-2R is consistent with the hypothesis that uORF translation is responsible for reduced expression of the main *NMT1* ORF.

Transcripts carrying mutations (Fig. 3a) at the first 5′ splice site (M14), the splicing branch site (M15) or the 3′ splice site (M16) result in uniformly low reporter expression (Fig. 3b, upper panel). RT–PCR analysis revealed that M14 yields I-2R RNA splicing product, whereas M15 and M16 do not undergo splicing (Fig. 3b, lower panel). These findings demonstrate that proper splicing is required to remove uORFs and permit main ORF expression.

For many bacterial riboswitches, metabolite binding alters folding of the expression platform located downstream of the aptamer without involving proteins[3,4,18]. To assess whether splicing regulation by the *NMT1* TPP riboswitch is due to protein-independent structural modulation of the aptamer flanks, we subjected *NMT1* UTR constructs to in-line probing[12]. Interestingly, the addition of TPP causes nucleotides at the branch site to become more structured (Supplementary Fig. 6), and yields a more flexible structure at the second 5′ splice site (Fig. 4a). Furthermore, we noticed that 12 nucleotides of the P4 and P5 elements of the aptamer are complementary to most of

the nucleotides at the second 5′ splice site that are structurally sequestered when ligand is absent (Fig. 4b). The P4 and P5 elements are required for recognition of the pyrophosphate moiety of TPP and, therefore, TPP binding and 5′ splice site occlusion are mutually exclusive.

The unusual characteristics of construct M9 in the *in vivo* reporter assays are consistent with this model for riboswitch function. In-line probing confirms that the M9 mutations disrupt base pairing between aptamer and the second 5′ splice site (Supplementary Fig. 7), and this structural defect is expected to favour the observed production of long spliced mRNA and the loss of reporter expression (Fig. 2c). Moreover, similar alternative base pairing potential exists for all TPP riboswitches associated with *NMT1* genes from other fungal species (Supplementary Fig. 8), indicating that this conserved alternative secondary structure is an important feature of the TPP riboswitch expression platform. It has been demonstrated that when presented with two 5′ splice sites, the spliceosome from the fungus *S. pombe* greatly prefers using the site proximal to the 3′ splice site[19]. Given this 5′ splice site preference, the TPP riboswitches in *NMT1* mRNAs can maintain complete control over the distribution of alternative splicing products simply by modulating base pairing between the P4–P5 aptamer region and the second 5′ splice site. TPP riboswitches in other fungal genes seem to use different mechanisms for gene control (Supplementary Fig. 9).

Our data are consistent with a mechanism for TPP riboswitch-mediated splicing regulation wherein metabolite binding alters the availability of alternative splice site and branch site components of the intron (Fig. 4c). When TPP concentration is low, the newly transcribed mRNA adopts a structure that occludes the second 5′ splice site, while leaving the branch site available for splicing. Pre-mRNA splicing from the first 5′ splice site leads to production of the I-3 form of mRNA and expression of the NMT1 protein. When TPP concentration is high, ligand binding to the TPP aptamer causes allosteric changes in RNA folding to increase the structural flexibility near the second 5′ splice site and to occlude nucleotides near the branch site. The combined effect of these changes is a reduction in splicing efficiency of the I-1 mRNA and a redirection of those I-1 mRNAs that do undergo processing to yield the alternatively spliced I-2 mRNA. Both I-1 and I-2 carry uORFs that compete with the translation of the main ORF and repress *NMT1* expression.

The involvement of alternative splicing in eukaryotic gene control is becoming increasingly apparent[20,21], and our findings reveal how riboswitches can modulate splicing efficiency and splice site choice without requiring protein factors. Given the enormous diversity of RNA folding possibilities, structured RNA domains are likely to be widely used to control splicing[22] through the direct read-out of physical changes such as temperature[23] or changes in metabolite concentrations[5,7,24]. Furthermore, an example of ligand-mediated control of

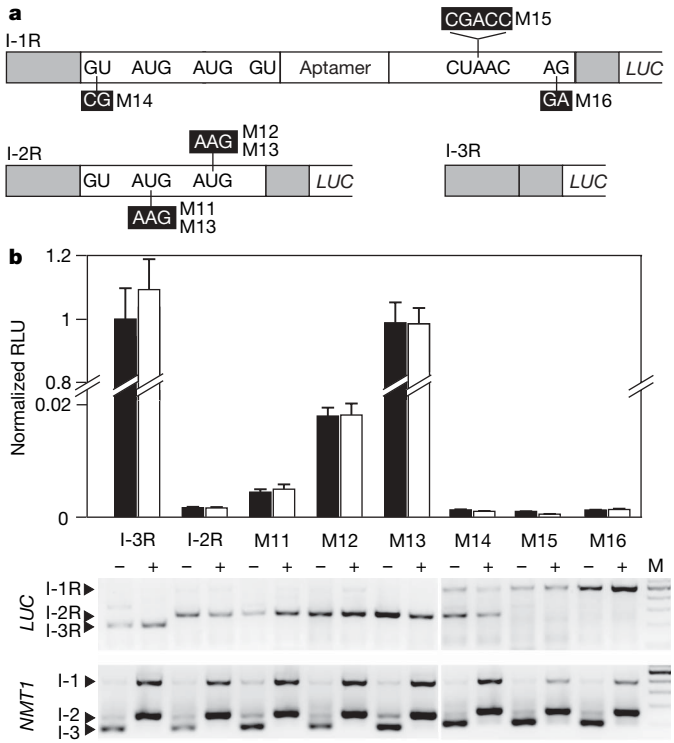


Figure 3 | Short uORFs in unspliced and alternatively spliced mRNAs cause *NMT1* repression. **a**, Wild-type and mutant constructs fused to a *LUC* reporter to simulate unspliced RNA (I-1R) and spliced RNAs (I-2R and I-3R). **b**, Upper panel, LUC activity in the absence (−, filled bars) or presence (+, open bars) of 30 μM thiamine in the medium. Expression was normalized relative to the value of the wild-type I-3R construct without addition of thiamine. Values are the averages from three independent assay repeats and standard deviation error bars are shown. Lower panel, RT–PCR analyses of the *LUC* fusion (upper panel) and native *NMT1* (lower panel) transcripts for *N. crassa* grown without (−) or with (+) thiamine. Details are described in the legend to Fig. 2c.

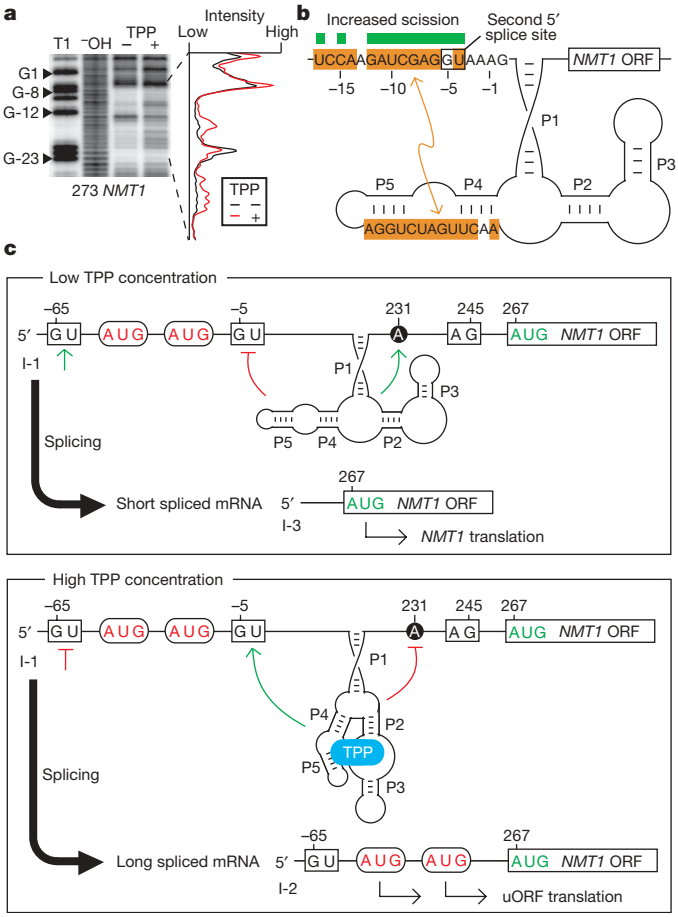


Figure 4 | Mechanism of TPP riboswitch-mediated alternative splicing of mRNA in *N. crassa*. **a**, TPP-induced modulation of structures near the second 5′ splice site. Spontaneous cleavage products of 5′ ^{32}P-labelled 273 *NMT1* RNA (nucleotides −78–195) were separated by polyacrylamide gel electrophoresis (PAGE) and quantified to reveal locations of 10 μM TPP-mediated changes in structure. T1 and ⁻OH designate partial RNA digestion with RNase T1 (cleaves at G residues) or alkali, respectively. Some RNase T1 product bands are labelled relative to the start of the aptamer as numbered in **b**. **b**, Some P4–P5 nucleotides are complementary (orange shading) to nucleotides near the second 5′ splice site that are modulated by TPP. **c**, Proposed mechanism for riboswitch control of *NMT1* expression in which key splicing determinants are activated (green lines) or inhibited (red lines) during different occupancy states of the aptamer.

splicing using an engineered aptamer has recently been reported[25], which demonstrates that direct ligand–mRNA interactions can be harnessed for gene control applications. Our observations with fungal TPP riboswitches further reveal the versatility of riboswitches from separate domains of life and hint at the possible involvement of undiscovered riboswitch classes in other gene control processes.

METHODS SUMMARY

Oligonucleotides and chemicals. RNAs were synthesized, synthetic DNAs (Supplementary Fig. 10) and reagents were purchased and DNA constructs were created, as noted in Methods.

RNA analyses. RT–PCR analyses were conducted using RNA from untransformed *N. crassa* inoculated into 100 ml of Vogel's minimal medium supplemented with 0.5 mg ml^{-1} L-histidine. Cultures were grown at 30 °C with shaking at 150 r.p.m. for 24 h either in the absence or presence of supplemented 30 μM thiamine. The complementary DNA was used as a template for PCR amplification of the 5′ regions of the three genes using primers as described in Supplementary Information. All splicing products were confirmed by cloning and sequencing.

Reporter gene assays. Transformation of *N. crassa* was conducted using electroporation of freshly suspended macroconidia and insertion of the target gene was verified by PCR with insert-specific primers from genomic DNA. Transcription of luciferase reporter constructs was constitutively driven by the *N. crassa* β-tubulin (*BTUB*) promoter inserted upstream of the *NMT1* 5′ UTR. *N. crassa* was grown overnight at 30 °C in 2% glucose minimal medium in the absence or presence of 30 μM thiamine. Samples were isolated and assayed for luciferase activity as described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 November 2006; accepted 20 March 2007.
Published online 29 April 2007.

1. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nature Rev. Mol. Cell Biol.* **5,** 451–463 (2004).
2. Winkler, W. C. & Breaker, R. R. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* **59,** 487–517 (2005).
3. Winkler, W. C., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956 (2002).
4. Mironov, A. S. *et al.* Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111,** 747–756 (2002).
5. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9,** 644–647 (2003).
6. Galagan, J. E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae. Nature* **438,** 1105–1115 (2005).
7. Kubodera, T. *et al.* Thiamine-regulated gene expression of *Aspergillus oryzae thiA* requires splicing of the intron containing a riboswitch-like domain in the 5′-UTR. *FEBS Lett.* **555,** 516–520 (2003).
8. McColl, D., Valencia, C. A. & Vierula, P. J. Characterization and expression of the *Neurospora crassa nmt-1* gene. *Curr. Genet.* **44,** 216–223 (2003).
9. Faou, P. & Tropschug, M. A novel binding protein for a member of CyP40-type Cyclophilins: *N. crassa* CyPBP37, a growth and thiamine regulated protein homolog to yeast Thi4p. *J. Mol. Biol.* **333,** 831–844 (2003).
10. Faou, P. & Tropschug, M. *Neurospora crassa* CyPBP37: a cytosolic stress protein that is able to replace yeast Thi4p function in the synthesis of vitamin B1. *J. Mol. Biol.* **344,** 1147–1157 (2004).
11. Maundrell, K. *nmt1* of fission yeast: a highly expressed gene completely repressed by thiamine. *J. Biol. Chem.* **265,** 10857–10864 (1989).
12. Soukup, G. A. & Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5,** 1308–1325 (1999).
13. Thore, S., Leibundgut, M. & Ban, N. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **312,** 1208–1211 (2006).
14. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441,** 1167–1171 (2006).
15. Edwards, T. E. & Ferré-D'Amaré, A. R. Crystal structures of the Thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA–small molecule recognition. *Structure* **14,** 1459–1468 (2006).
16. Welz, R. & Breaker, R. R. Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis. RNA* advance online publication (16 February, doi:10.1261/rna.407707 (2007).
17. Vilela, C. & McCarthy, J. E. Regulation of fungal gene expression via short open reading frames in the mRNA 5′ untranslated region. *Mol. Microbiol.* **49,** 859–867 (2003).
18. Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9,** 1043–1049 (2002).
19. Romfo, C. M., Alvarez, C. J., van Heeckeren, W. J., Webb, C. J. & Wise, J. A. Evidence for splice site pairing via intron definition in *Schizosaccharomyces pombe. Mol. Cell. Biol.* **20,** 7955–7970 (2000).
20. Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nature Rev. Mol. Cell Biol.* **6,** 386–398 (2005).
21. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126,** 37–47 (2006).
22. Buratti, E. & Baralle, F. E. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.* **24,** 10505–10514 (2004).
23. Colot, H. V., Loros, J. J. & Dunlap, J. C. Temperature-modulated alternative splicing and promoter use in the circadian clock gene frequency. *Mol. Biol. Cell* **16,** 5563–5571 (2005).
24. Borsuk, P. *et al.* L-Arginine influences the structure and function of arginase mRNA in *Aspergillus nidulans. Biol. Chem.* **388,** 135–144 (2007).
25. Kim, D.-S., Gusti, V., Pillai, S. G. & Gaur, R. K. An artificial riboswitch for controlling pre-mRNA splicing. *RNA* **11,** 1667–1677 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. We thank J. C. Dunlap for hosting and training M.T.C. in fungal genetics techniques and for supplying us with *N. crassa* strains, genomic DNA of *N. crassa* and the luciferase reporter vector. Also, we thank members of the Dunlap laboratory (M. Shi and L. Larrondo) for technical assistance and helpful discussions. A.W. was supported by a postdoctoral fellowship from the German Research Foundation (DFG) and M.T.C. was supported by a Yale College Dean's Fellowship and an HHMI Future Scientist Fellowship. This work also was supported by an NIH grant (R.R.B).

Author Contributions M.T.C. and A.W. conducted all genetic and molecular biology analyses, and M.T.C. and N.S. conducted biochemical analyses. M.T.C., A.W. and R.R.B. co-wrote the manuscript. All authors contributed to discussions regarding the data and their interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to R.R.B. (ronald.breaker@yale.edu).

METHODS

Bioinformatics searches and fungal TPP riboswitches. We examined the 'fungi' division of the RefSeq database (version 13) using covariance model searches with manually curated seed sequence alignments adapted from known TPP riboswitch representatives. Covariance models[26] were created using the INFERNAL software package[27] (version 0.55). See also Supplementary Information for additional details.

DNA oligonucleotides and chemicals. Synthetic DNAs were purchased from the HHMI Keck Foundation Biotechnology Resource Center at Yale University. TPP, thiamine, sodium iodoacetate (IAA) and L-histidine were purchased from Sigma-Aldrich. [γ-^{32}P]ATP was purchased from Amersham Pharmacia.

***In vitro* transcription.** DNA templates were produced by PCR amplification from genomic DNA of *N. crassa* using primers designed to introduce a T7 promoter into the construct. The sequence CC was added to the template strand transcription start site to promote efficient *in vitro* transcription, thus producing RNAs that carry GG at their 5′ terminus. RNAs were prepared using a RiboMax Transcription Kit (Promega) according to the manufacturer's directions. RNAs were purified by denaturing polyacrylamide gel electrophoresis (PAGE), and 5′ ^{32}P-labelled as described previously[28].

In-line probing of RNA constructs. 5′ ^{32}P-labelled RNAs were incubated at 23 °C for 40 h in 50 mM Tris-HCl (pH 8.3 at 25 °C), 20 mM $MgCl_2$ and 100 mM KCl in the presence or absence of TPP as defined. Cleavage products were separated by denaturing 10% PAGE, visualized by PhosphorImager (GE Healthcare), and quantified using ImageQuant software. K_D values were determined by plotting the normalized fraction of RNA cleaved versus the logarithm of ligand concentration used.

Strains, plasmids and media. *N. crassa* 87–74 (*bd*; *frq*$^{+}$ *a*; *his-3*)[29] was used as a host strain for transformation. The plasmid pLL07 (provided by the laboratory of J. C. Dunlap)[30], which carries a firefly luciferase (*LUC*) reporter gene, was used for reporter gene construction. The start codon for the *LUC* reporter gene in pLL07 was removed by QuikChange (Stratagene) site directed mutagenesis to obtain the plasmid pLL09.

The promoter for the β-tubulin gene in *N. crassa* ranging from positions 1 to 355 (accession number M13630)[31] was amplified from genomic DNA by PCR using primers DNA3 and DNA4. The amplified DNA fragment was digested with *Mfe*I and *Eco*RI and inserted into the *Eco*RI site of pLL09 to obtain pLUC. The sequence of pLUC was confirmed by sequencing (HHMI Keck Foundation Biotechnology Resource Center at Yale University). *E. coli* Top 10 cells (Invitrogen) were used as a host during manipulation of plasmids.

Cloning of the 5′ UTR of the *NMT1* gene (accession number AY007661) was achieved by PCR amplification of a 378 base pair fragment (beginning with the annotated transcription start site) from genomic DNA of *N. crassa* with primers DNA5 and DNA6. The resulting wild-type (WT) PCR DNA was first cloned using a TOPO TA cloning kit (Invitrogen). The *NMT1* fragment was released from the vector by *Eco*RI and *Xba*I restriction enzyme digestion and cloned into appropriate sites of pLUC.

For generation of the aptamer mutants M1 through M9 and the splice site mutants M14 through M16, PCR mutagenesis was performed on the WT *NMT1*-containing TOPO vector (see primer list). After confirmation of mutagenesis by DNA sequencing, each mutant *NMT1* fragment was cloned into *Eco*RI/*Xba*I sites of pLUC. For cloning of the P3 deletion construct (M10), *NMT1* was amplified in two fragments with primers DNA5/DNA25 and DNA26/DNA6, in which the overlapping region deletes much of the natural P3a stem. These two fragments were used as a template in a subsequent PCR with the outer primers DNA5 and DNA6. The resulting fragment was digested with *Eco*RI and *Xba*I and cloned into pLUC. Preparation of the I-2R and I-3R constructs and variants of *NMT1* was achieved by RT–PCR amplification of these two alternatively spliced products with primers DNA5 and DNA6. The resulting PCR products were cloned and mutated as described above.

To generate a *NCU01977.1*-5′ region–*LUC* reporter fusion (Supplementary Fig. 9), a 478 nucleotide fragment starting 94 nucleotides upstream of the predicted start codon was amplified from genomic DNA of *N. crassa* with primers DNA41 and DNA42 and cloned into *Eco*RI/*Xba*I sites of pLUC as described above. The integrity of all constructs was confirmed by sequencing (HHMI Keck Foundation Biotechnology Resource Center at Yale University). In all constructs, the original start codon of the main ORF (*NMT1* or *NCU01977.1*) was fused in frame with the *LUC* reporter sequence.

Standard liquid medium used for growth of *N. crassa* contained 2% glucose, 0.5% L-arginine, 1× Vogel's minimal medium, and 50 ng ml^{-1} biotin. Solid medium used for *N. crassa* growth (slants) contained 1× Vogel's minimal medium, 2% sucrose and 1.5% agar. Medium used for selection of *N. crassa* transformants contained 1× Vogel's minimal medium, 2% agar, 2% L-sorbose, 0.05% fructose and 0.05% glucose. For homokaryon isolation, 0.1× Westergaard's medium containing 1% IAA was used[32].

***N. crassa* transformations and reporter assays.** Transformation of *N. crassa* was conducted using electroporation of freshly suspended macroconidia as previously described[33–35]. Insertion of the target gene was verified by PCR with insert specific primers from genomic DNA. Homokaryotic strains were isolated as previously described[36].

Luciferase reporter gene assays. Mycelia from *N. crassa* transformed with LUC reporter constructs were isolated by filtration and approximately 100 mg of tissue was ground to a fine powder. After addition of 100 µl 1× Passive Lysis Buffer (Promega), the samples were vigorously mixed and incubated on ice for 30 min followed by centrifugation for 15 min at 13,000 × *g*. Luciferase activity was determined for the resulting supernatant using the Luciferase Assay System (Promega) and a plate-reading luminometer (Wallac). Luciferase activity was normalized over total protein concentration of the extract as determined by Bradford Protein Assay (BioRad) and finally expressed relative to a reference construct. Luciferase background activity (untransformed *N. crassa*) was 0.05% relative to the wild-type *NMT1* construct in cells grown without added thiamine.

Reverse transcriptase–polymerase chain reaction (RT–PCR) analyses. Total RNA was isolated from mycelia using TRIzol LS reagent (Invitrogen) according to the manufacturer's directions. Five µg of total RNA was treated with RNase free DNase I (Promega) for 30 min at 37 °C. Complementary DNA was generated by reverse transcription with a polyT primer for 1 h at 42 °C using SuperScript II Reverse Transcriptase (Invitrogen) according to the manufacturer's directions. To exclude the possibility of amplification products originating from contamination with genomic DNA, control reactions using RNA preparations before reverse transcription were performed and, for *NMT1*, a reverse primer spanning an exon–exon border in the coding region was used. Additional details regarding RT–PCR analyses are provided in Supplementary Information.

26. Eddy, S. R. & Durbin, R. RNA sequence analysis using covariance models. *Nucleic Acids Res.* **22,** 2079–2088 (1994).
27. Eddy, S. R. INFERNAL. Version 0.55. Distributed by the author. Department of Genetics, Washington Univ. School of Medicine, St. Louis, Missouri.
28. Seetharaman, S., Zivarts, M., Sudarsan, N. & Breaker, R. R. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nature Biotechnol.* **19,** 336–341 (2001).
29. Froehlich, A. C., Loros, J. J. & Dunlap, J. C. Rhythmic binding of a WHITE COLLAR-containing complex to the frequency promoter is inhibited by FREQUENCY. *Proc. Natl Acad. Sci. USA* **100,** 5914–5919 (2003).
30. Mehra, A., Morgan, L., Bell-Pedersen, D., Loros, J. & Dunlap, J. C. Watching the *Neurospora* clock tick. *Soc. Res. Biol. Rhythms* 27 (Amelia Island, Florida, Society for Research on Biological Rhythms, 22–25 May, 2002).
31. Orbach, M. J., Porro, E. B. & Yanofsky, C. Cloning and characterization of the gene for β-tubulin from a benomyl-resistant mutant of *Neurospora crassa* and its use as a dominant selectable marker. *Mol. Cell. Biol.* **6,** 2452–2461 (1986).
32. Westergaard, M. & Mitchell, H. K. *Neurospora* V. A synthetic medium favoring sexual reproduction. *Am. J. Bot.* **34,** 573–577 (1947).
33. Davis, R. H. Neurospora: *Contributions of a Model Organism*. (Oxford Univ. Press, New York, New York, 2000).
34. Loros, J. J. & Dunlap, J. C. *Neurospora crassa* clock-controlled genes are regulated at the level of transcription. *Mol. Cell. Biol.* **11,** 558–563 (1991).
35. Vann, D. C. Electroporation-based transformation of freshly harvested conidia of *Neurospora crassa. Fungal Genet. Newsl.* **42A,** 53 (1995).
36. Ebbole, D. & Sachs, M. S. A rapid and simple method for isolation of *Neurospora crassa* homokaryons using microconidia. *Fungal Genet. Newsl.* **37,** 17–18 (1990).

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

Americans of a certain age may remember a popular Saturday morning television programme called *Super Friends*. This cartoon series featured, among other characters, the Wonder Twins. Whenever these two superheroes got into a jam, they touched their knuckles together and unleashed their powers, transforming themselves into other shapes and forms — to the detriment of the bad guy, who they quickly vanquished.

That coupling reminds me of the two largest global foundations that fund postdocs. The Howard Hughes Medical Institute (HHMI) and the Wellcome Trust each stand alone as powerful, influential organizations, funding some of the best and brightest investigators and postdocs globally. But last week, they metaphorically touched their knuckles together, launching a programme that would allow fellows funded by either programme to spend a year in a lab of their transatlantic allies.

This postdoctoral exchange programme should be a boon to postdocs funded by either organization because it will expose them to potential future collaborators on both sides of the ocean and allow them to learn different skills — furthering interdisciplinary science. They will also get to meet and work with investigators who are not necessarily funded by either foundation. This arrangement broadens career paths for participating postdocs, by making them less beholden to any particular principal investigator or even single laboratory.

But the real test of how 'super' this programme will be is whether other organizations and institutes take up this model. Both the HHMI and the Wellcome Trust have set the pace in postdoc training by raising stipend levels and offering courses beyond scientific skills, such as in lab management. In both cases, many other organizations followed suit. It's likely that, as with the Wonder Twins, the whole will be greater than the sum of their parts, in ways I can't envisage — much as I still can't comprehend how two kids in capes changing into an eagle and a bucket of water can somehow prevail over their nefarious adversaries.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

L. GILLIERON/CORBIS

Gates of opportunity

Seattle is reaping the benefits of having the world's largest health foundation in its backyard. **Eric Sorensen** gauges the impact.

The Bill & Melinda Gates Foundation aims to address the most pressing public-health issues around the planet. And in its search for solutions, especially to diseases affecting the poorest people, the world's wealthiest foundation is spending a lot of money close to its Seattle home.

The Fred Hutchinson Cancer Research Center, which last year received $40 million in Gates funding for work on an HIV vaccine, is within walking distance of the foundation's office on Eastlake Avenue. The University of Washington, home of a new global-health department started with $30 million in Gates funding and recipient of another $10 million for AIDS vaccine work, is 10 minutes away on the 70 bus.

The non-profit Program for Appropriate Technology in Health (PATH) has received $850 million in Gates funding over a dozen years. Its new and already bulging office building is in Seattle's Ballard neighbourhood, a kayak paddle down Lake Union's ship canal. The Seattle Biomedical Research Institute (SBRI) has received Gates grants totalling nearly $45 million. Its benefactor will soon be even closer: in 2010, the Gates Foundation is due to move into a new 56,000-square-metre headquarters close to the nearby Space Needle tower.

The local funding recognizes Seattle's growing research prowess, particularly in health problems such as malaria, tuberculosis and HIV/AIDS, and in innovative technological solutions. More than half the Gates donations go into global health, with 15% spent in the Seattle area. Global-health researchers around Seattle have received more than $1 billion since Microsoft co-founder Bill Gates and his wife Melinda started the foundation in 1994. And investment in the region is likely to increase. A $30-billion pledge from investor Warren Buffett last year doubled the foundation's size, so staffing will follow. As the foundation grows, it plans to make its involvement in its current issues "deeper and not broader", according to Melinda Gates.

"The effect of the Buffett gift is now beginning to be felt," says Jack Faris, a former spokesman for the foundation who is now president of the Washington Biotechnology and Biomedical Association.

PATH

Chris Elias: collaboration inspired by Northwest spirit.

Key collaborations

"The Seattle area has the capacity to highly effectively collaborate worldwide on complex, important contemporary problems and projects," says Jim Gore, the SBRI's chief operating officer. "Every position, from leadership to all our scientific career levels, we expect to see expand locally and we expect to stimulate growth through our collaborators. I don't think there is a job classification that will be left behind."

Faris and others say the foundation's tapping of local expertise is anything but parochial, as global-health research has been part of the Seattle fabric for decades. The SBRI began studying malaria parasites here 30 years ago and PATH has been around nearly as long. The 'Hutch', as the Hutchinson Center is often called, is a leader not only in cancer research but also in the study of HIV/AIDS and other diseases that compromise the immune system. This has led the Hutch to become the coordinating site for many

FRED HUTCHINSON RESEARCH CENTER

The Hutch: seeking innovative treatments for cancer and other diseases.

national and international studies, such as the Women's Health Initiative and the HIV Vaccine Trials Network. The University of Washington is also a notable player, housing the Center for AIDS and STD since 1989.

The region's life-sciences and health researchers have a history of collaboration. This may stem from the city's distance from competitive places such as New York, says Chris Elias, PATH president, or maybe it's what he calls the "Northwest spirit". Whatever the cause, when researchers from the university and from a local biotechnology company needed a Biosafety Level 3 facility, the SBRI provided space and time. "That's the kind of sharing they would not be doing in many places that are hotly competing," says Elias.

Developing solutions

The rapidly increasing number of biotech firms and researchers bring other benefits too. They can collectively push for key policies such as better public education, and the availability of jobs makes it easier to hire researchers with spouses who also work. "The bigger the community gets, the easier it gets to recruit," Elias says.

For PATH, this means high-tech solutions for the low-tech developing world with its poor transport, patchy refrigeration and weak health infrastructure. Not only does it work on low-cost vaccines for malaria and meningitis, it is also working on ways to improve access to vaccines for other conditions, such as hepatitis B.

Until about five years ago, there was usually about a 20-year gap between the United States and the developing world getting access to a vaccine, says Teresa Guillien, a spokeswoman for PATH, which is trying to close that gap. In 2005, PATH got a Gates grant of $107.6 million to work with GlaxoSmithKline Biologicals to complete testing and licensing of the most advanced potential vaccine for malaria.

Gates has also helped fund several products at PATH including single-use, self-disabling syringes, an inexpensive dipstick to test for HIV antibodies and vitamin-fortified rice. A PATH group, led by bioengineer Paul Yager of the University of Washington and working with a local diagnostics company called Micronics as well as other firms, is helping develop a credit-card-sized micro lab that can quickly diagnose blood or stool samples for diseases that cause fever. One of two PATH 'lab on a card' projects, it was among 43 chosen from 1,500 responses to the foundation's Grand Challenges to improve global health in 2003. PATH officials see numerous opportunities for multidisciplinary projects that involve technical work, commercialization, public-health expertise and even industrial design.

Stefan Kappe is working on ways to fight malaria.

In Seattle's South Lake Union area, the SBRI has two recipients of Grand Challenges grants totalling $32.5 million. Parasitologist Stefan Kappe is working on mosquitoes to genetically attenuate the *Plasmodium falciparum* parasite, which spreads malaria. Patrick Duffy is researching children's immune responses, to learn why some suffer so much more severely than others from malaria. He is also working on a vaccine to block the protein that helps the parasite bond to the placenta and rob the fetus of nutrition.

Rapid expansion

"The number and scope of laboratories in the Pacific Northwest conducting research on global health has expanded substantially over the past five years, and will continue to do so in the near term," says Duffy, crediting Gates support and the other funding it has catalysed. "The Gates Foundation itself seems to be on a growth trajectory so there may be new opportunities in analysis, policy and programme management at the foundation."

Duffy has seen the SBRI branch out, with its project managers helping to translate discoveries into products and providing opportunities for people who achieved success in biotech companies and now want to go into non-profit areas. "This is a great opportunity for them in the Seattle area and elsewhere," Duffy says.

SBRI

Julie McElrath, a Hutchinson researcher, is lead investigator on a $30-million grant to study ways to enhance the cellular immune response generated by HIV vaccines. She talks about the need for scientists who have moved into business, earning either a law degree or MBA, and has seen many staff scientists come from industry. "They understand what it takes to develop a product better than a standard research scientist," she says. "They understand how to work with milestones."

Complex skill sets are needed to meet the Gates Foundation's requirement that researchers share data and collaborate in real time. Its new $287-million HIV/AIDS consortium, for example, has 165 researchers working in 16 teams. Analysing much of its data is Steven Self, head of the Hutchinson Center's Statistical Center for HIV/AIDS Research and Prevention. Self and his colleagues are leading a project to create a repository of statistical data on vaccine candidates being tested within the research network.

The network is likely to produce a wealth of connections and collaborations "that otherwise might never be made", he says. This creates opportunities for database statisticians, mathematical modellers thinking about dynamical systems, and statisticians focused on complex, multidimensional immunological data.

"There are not just more positions of the usual sort," says Self. "There is a wider variety of positions that we're looking for, to solve a wider variety of problems." These are some of the biggest problems in the world and, with support from the Gates Foundation, Seattle is helping to look for innovative solutions. ■

Eric Sorensen is a science writer based in Seattle.

MOVERS

Glenn Morris, director, Emerging Pathogens Institute, University of Florida, Gainesville, Florida

2005–07: Interim dean, University of Maryland (UM) School of Public Health, Baltimore, Maryland
2000–07: Chairman, Department of Epidemiology and Preventive Medicine, UM School of Medicine
1996–2000: Head, Division of Hospital Epidemiology Department of Medicine, UM School of Medicine

As a child of missionary parents based in Bangkok, Thailand, Glenn Morris saw at first hand the waves of pandemic cholera that struck southeast Asia each summer. "That experience left me with the indelible sense of the drama and urgency of dealing with an emerging pathogen," he says.

Back in the United States for college, Morris got a bachelor's degree in both biology and history from Rice University in Houston, Texas. He then went on to take an MD at Tulane University School of Medicine in New Orleans, Louisiana. At Tulane, he also studied dengue virus while getting his master's in public health and tropical medicine.

Later, as an epidemic intelligence officer with the US Centers for Disease Control and Prevention (CDC), he dealt with cholera outbreaks at refugee camps in Thailand. Upon his return to the United States, Morris completed his residency before beginning a fellowship — and what turned out to be a 25-year career — at the Center for Vaccine Development at the University of Maryland. Focusing on the molecular genetics of pathogens, he developed molecular markers and used molecular fingerprinting to understand transmission pathways.

Morris honed his laboratory techniques over the next decade, then used them in clinical epidemiological research. He also served on National Academy of Sciences panels, addressing government policies on food-borne disease. An outspoken critic, he helped restructure US Department of Agriculture regulations to include the first microbial standards for food safety. "That was an exciting couple of years, watching ideas take shape into regulation," he says.

He now plans to use his public health and policy know-how as the director of the University of Florida's new $50-million Emerging Pathogens Institute in Gainesville. Myron Levine, director of the University of Maryland's Center for Vaccine Development, says that Morris's wide-ranging mix of talents make him well suited to head an institute in a state susceptible to myriad food- and water-borne infectious diseases.

Noting Morris's work in Thailand, James Hughes, a former director of the CDC's centre for infectious diseases who is now at Emory University in Atlanta, Georgia, says that although many in the infectious diseases field try to combine strengths in epidemiology and microbiology, "solid international experience is vitally important". ■

Virginia Gewin

BRICKS & MORTAR

Northern exposure for EMBL

The European Molecular Biology Laboratory (EMBL), Europe's foremost molecular-biology institution, is extending its reach to the north. As part of efforts to establish laboratory partnerships with Nordic countries, EMBL plans this year to set up a new laboratory in Umeå, a coastal city on the Gulf of Bothnia, 650 kilometres north of Stockholm. One of the hopes, says future laboratory director Bernt Eric Uhlin, is to build a sustainable national resource that will link Europe's Nordic countries to EMBL, which currently has five labs in Germany, France, Italy and Britain.

The lab will be part of the first EMBL partnership in molecular infectious medicine in the Nordic countries. Others will include the Institute for Molecular Medicine Finland, which specializes in genetic epidemiology, and the Centre for Molecular Biology and Neuroscience at the University of Oslo.

The goal is for Umeå to become a Swedish node in EMBL. The new funding will allow up to seven young researchers to be recruited internationally and to establish research groups, for the time being in existing buildings and departments.

The new laboratory will focus on molecular infectious medicine, drawing from 25 existing affiliated groups in microbiology, molecular biology, chemistry and physics that collaborate via the Umeå Centre for Microbial Research. Known as one of the best places to study bacterial pathogenesis in Europe, the centre has expertise in infection biology and molecular microbiology, says Uhlin. "The new groups should thrive in this scientific environment," he says.

The partnership gives scientists the chance to work for six to eight years "to develop their skills in an interesting environment and then go on to other universities", says Lars Börjesson, secretary-general of the Swedish Research Council's Committee for Research Infrastructures.

It also stands to draw international attention to Sweden's biomedical research. "We think there should be more European exposure of Sweden's research," says Börjesson, who hopes it will also stimulate Sweden's biomedical research community.

The Swedish government provided just under half of the funding, with the rest coming from Umeå University. The total invested in the project is €18.5 million (US$25 million). Most will be spent on staff, but some will be invested in shared technical platforms and equipment. ■

Hannah Hoag

POSTDOC JOURNAL

Perception versus reality

The closer the meeting got, the more I dreaded going. I hadn't had much interest in my work before the meeting, and I also hated going so far from home. I'm in my third trimester of pregnancy, so I am very big and I tire easily. The idea of schlepping my luggage all over airports and sleeping in a strange bed was not appealing. But I went.

The travel was terrible; we missed three flights, dinner and both plenary talks on the first day. Once there, things improved. I chatted with many other pregnant women in the community about how they're coping. There was even a husband and wife who took turns to care for their 18-month-old son while the other parent attended individual sessions. When I spoke to the husband, he asked why I hadn't brought my son along.

I have to admit, I was surprised to find out how many researchers, men and women, had small children. In general, other researchers' family lives seemed so hidden to me before now. Are such discussions taboo or just mundane? I used to think it was the former, but now I'm inclined to believe it's the latter. For me, it's just life. I often feel as if I have to explain to others what having children entails — especially to single people and those without children. But now I see that plenty of other people don't think it's an extraordinary feat. Who was I trying to convince, them or myself? ■

Moira Sheehan is a postdoc in plant breeding and genetics at Cornell University.

MOVERS

Glenn Morris, director, Emerging Pathogens Institute, University of Florida, Gainesville, Florida

2005–07: Interim dean, University of Maryland (UM) School of Public Health, Baltimore, Maryland
2000–07: Chairman, Department of Epidemiology and Preventive Medicine, UM School of Medicine
1996–2000: Head, Division of Hospital Epidemiology Department of Medicine, UM School of Medicine

As a child of missionary parents based in Bangkok, Thailand, Glenn Morris saw at first hand the waves of pandemic cholera that struck southeast Asia each summer. "That experience left me with the indelible sense of the drama and urgency of dealing with an emerging pathogen," he says.

Back in the United States for college, Morris got a bachelor's degree in both biology and history from Rice University in Houston, Texas. He then went on to take an MD at Tulane University School of Medicine in New Orleans, Louisiana. At Tulane, he also studied dengue virus while getting his master's in public health and tropical medicine.

Later, as an epidemic intelligence officer with the US Centers for Disease Control and Prevention (CDC), he dealt with cholera outbreaks at refugee camps in Thailand. Upon his return to the United States, Morris completed his residency before beginning a fellowship — and what turned out to be a 25-year career — at the Center for Vaccine Development at the University of Maryland. Focusing on the molecular genetics of pathogens, he developed molecular markers and used molecular fingerprinting to understand transmission pathways.

Morris honed his laboratory techniques over the next decade, then used them in clinical epidemiological research. He also served on National Academy of Sciences panels, addressing government policies on food-borne disease. An outspoken critic, he helped restructure US Department of Agriculture regulations to include the first microbial standards for food safety. "That was an exciting couple of years, watching ideas take shape into regulation," he says.

He now plans to use his public health and policy know-how as the director of the University of Florida's new $50-million Emerging Pathogens Institute in Gainesville. Myron Levine, director of the University of Maryland's Center for Vaccine Development, says that Morris's wide-ranging mix of talents make him well suited to head an institute in a state susceptible to myriad food- and water-borne infectious diseases.

Noting Morris's work in Thailand, James Hughes, a former director of the CDC's centre for infectious diseases who is now at Emory University in Atlanta, Georgia, says that although many in the infectious diseases field try to combine strengths in epidemiology and microbiology, "solid international experience is vitally important". ■

Virginia Gewin

BRICKS & MORTAR

Northern exposure for EMBL

The European Molecular Biology Laboratory (EMBL), Europe's foremost molecular-biology institution, is extending its reach to the north. As part of efforts to establish laboratory partnerships with Nordic countries, EMBL plans this year to set up a new laboratory in Umeå, a coastal city on the Gulf of Bothnia, 650 kilometres north of Stockholm. One of the hopes, says future laboratory director Bernt Eric Uhlin, is to build a sustainable national resource that will link Europe's Nordic countries to EMBL, which currently has five labs in Germany, France, Italy and Britain.

The lab will be part of the first EMBL partnership in molecular infectious medicine in the Nordic countries. Others will include the Institute for Molecular Medicine Finland, which specializes in genetic epidemiology, and the Centre for Molecular Biology and Neuroscience at the University of Oslo.

The goal is for Umeå to become a Swedish node in EMBL. The new funding will allow up to seven young researchers to be recruited internationally and to establish research groups, for the time being in existing buildings and departments.

The new laboratory will focus on molecular infectious medicine, drawing from 25 existing affiliated groups in microbiology, molecular biology, chemistry and physics that collaborate via the Umeå Centre for Microbial Research. Known as one of the best places to study bacterial pathogenesis in Europe, the centre has expertise in infection biology and molecular microbiology, says Uhlin. "The new groups should thrive in this scientific environment," he says.

The partnership gives scientists the chance to work for six to eight years "to develop their skills in an interesting environment and then go on to other universities", says Lars Börjesson, secretary-general of the Swedish Research Council's Committee for Research Infrastructures.

It also stands to draw international attention to Sweden's biomedical research. "We think there should be more European exposure of Sweden's research," says Börjesson, who hopes it will also stimulate Sweden's biomedical research community.

The Swedish government provided just under half of the funding, with the rest coming from Umeå University. The total invested in the project is €18.5 million (US$25 million). Most will be spent on staff, but some will be invested in shared technical platforms and equipment. ■

Hannah Hoag

POSTDOC JOURNAL

Perception versus reality

The closer the meeting got, the more I dreaded going. I hadn't had much interest in my work before the meeting, and I also hated going so far from home. I'm in my third trimester of pregnancy, so I am very big and I tire easily. The idea of schlepping my luggage all over airports and sleeping in a strange bed was not appealing. But I went.

The travel was terrible; we missed three flights, dinner and both plenary talks on the first day. Once there, things improved. I chatted with many other pregnant women in the community about how they're coping. There was even a husband and wife who took turns to care for their 18-month-old son while the other parent attended individual sessions. When I spoke to the husband, he asked why I hadn't brought my son along.

I have to admit, I was surprised to find out how many researchers, men and women, had small children. In general, other researchers' family lives seemed so hidden to me before now. Are such discussions taboo or just mundane? I used to think it was the former, but now I'm inclined to believe it's the latter. For me, it's just life. I often feel as if I have to explain to others what having children entails — especially to single people and those without children. But now I see that plenty of other people don't think it's an extraordinary feat. Who was I trying to convince, them or myself? ■

Moira Sheehan is a postdoc in plant breeding and genetics at Cornell University.

MOVERS

Glenn Morris, director, Emerging Pathogens Institute, University of Florida, Gainesville, Florida

2005–07: Interim dean, University of Maryland (UM) School of Public Health, Baltimore, Maryland
2000–07: Chairman, Department of Epidemiology and Preventive Medicine, UM School of Medicine
1996–2000: Head, Division of Hospital Epidemiology Department of Medicine, UM School of Medicine

As a child of missionary parents based in Bangkok, Thailand, Glenn Morris saw at first hand the waves of pandemic cholera that struck southeast Asia each summer. "That experience left me with the indelible sense of the drama and urgency of dealing with an emerging pathogen," he says.

Back in the United States for college, Morris got a bachelor's degree in both biology and history from Rice University in Houston, Texas. He then went on to take an MD at Tulane University School of Medicine in New Orleans, Louisiana. At Tulane, he also studied dengue virus while getting his master's in public health and tropical medicine.

Later, as an epidemic intelligence officer with the US Centers for Disease Control and Prevention (CDC), he dealt with cholera outbreaks at refugee camps in Thailand. Upon his return to the United States, Morris completed his residency before beginning a fellowship — and what turned out to be a 25-year career — at the Center for Vaccine Development at the University of Maryland. Focusing on the molecular genetics of pathogens, he developed molecular markers and used molecular fingerprinting to understand transmission pathways.

Morris honed his laboratory techniques over the next decade, then used them in clinical epidemiological research. He also served on National Academy of Sciences panels, addressing government policies on food-borne disease. An outspoken critic, he helped restructure US Department of Agriculture regulations to include the first microbial standards for food safety. "That was an exciting couple of years, watching ideas take shape into regulation," he says.

He now plans to use his public health and policy know-how as the director of the University of Florida's new $50-million Emerging Pathogens Institute in Gainesville. Myron Levine, director of the University of Maryland's Center for Vaccine Development, says that Morris's wide-ranging mix of talents make him well suited to head an institute in a state susceptible to myriad food- and water-borne infectious diseases.

Noting Morris's work in Thailand, James Hughes, a former director of the CDC's centre for infectious diseases who is now at Emory University in Atlanta, Georgia, says that although many in the infectious diseases field try to combine strengths in epidemiology and microbiology, "solid international experience is vitally important". ■

Virginia Gewin

BRICKS & MORTAR

Northern exposure for EMBL

The European Molecular Biology Laboratory (EMBL), Europe's foremost molecular-biology institution, is extending its reach to the north. As part of efforts to establish laboratory partnerships with Nordic countries, EMBL plans this year to set up a new laboratory in Umeå, a coastal city on the Gulf of Bothnia, 650 kilometres north of Stockholm. One of the hopes, says future laboratory director Bernt Eric Uhlin, is to build a sustainable national resource that will link Europe's Nordic countries to EMBL, which currently has five labs in Germany, France, Italy and Britain.

The lab will be part of the first EMBL partnership in molecular infectious medicine in the Nordic countries. Others will include the Institute for Molecular Medicine Finland, which specializes in genetic epidemiology, and the Centre for Molecular Biology and Neuroscience at the University of Oslo.

The goal is for Umeå to become a Swedish node in EMBL. The new funding will allow up to seven young researchers to be recruited internationally and to establish research groups, for the time being in existing buildings and departments.

The new laboratory will focus on molecular infectious medicine, drawing from 25 existing affiliated groups in microbiology, molecular biology, chemistry and physics that collaborate via the Umeå Centre for Microbial Research. Known as one of the best places to study bacterial pathogenesis in Europe, the centre has expertise in infection biology and molecular microbiology, says Uhlin. "The new groups should thrive in this scientific environment," he says.

The partnership gives scientists the chance to work for six to eight years "to develop their skills in an interesting environment and then go on to other universities", says Lars Börjesson, secretary-general of the Swedish Research Council's Committee for Research Infrastructures.

It also stands to draw international attention to Sweden's biomedical research. "We think there should be more European exposure of Sweden's research," says Börjesson, who hopes it will also stimulate Sweden's biomedical research community.

The Swedish government provided just under half of the funding, with the rest coming from Umeå University. The total invested in the project is €18.5 million (US$25 million). Most will be spent on staff, but some will be invested in shared technical platforms and equipment. ■

Hannah Hoag

POSTDOC JOURNAL

Perception versus reality

The closer the meeting got, the more I dreaded going. I hadn't had much interest in my work before the meeting, and I also hated going so far from home. I'm in my third trimester of pregnancy, so I am very big and I tire easily. The idea of schlepping my luggage all over airports and sleeping in a strange bed was not appealing. But I went.

The travel was terrible; we missed three flights, dinner and both plenary talks on the first day. Once there, things improved. I chatted with many other pregnant women in the community about how they're coping. There was even a husband and wife who took turns to care for their 18-month-old son while the other parent attended individual sessions. When I spoke to the husband, he asked why I hadn't brought my son along.

I have to admit, I was surprised to find out how many researchers, men and women, had small children. In general, other researchers' family lives seemed so hidden to me before now. Are such discussions taboo or just mundane? I used to think it was the former, but now I'm inclined to believe it's the latter. For me, it's just life. I often feel as if I have to explain to others what having children entails — especially to single people and those without children. But now I see that plenty of other people don't think it's an extraordinary feat. Who was I trying to convince, them or myself? ■

Moira Sheehan is a postdoc in plant breeding and genetics at Cornell University.